

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**Desenvolvimento de Modelos de Regressão Logística para a
avaliação de novas tecnologias de mobilidade urbana**

Joana Alexandre de Vasconcelos Dias

Trabalho de Projeto
Mestrado em Estatística e Investigação Operacional
Especialização em Estatística

2015

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**Desenvolvimento de Modelos de Regressão Logística para a
avaliação de novas tecnologias de mobilidade urbana**

Joana Alexandre de Vasconcelos Dias

Trabalho de Projeto

Mestrado em Estatística e Investigação Operacional

Especialização em Estatística

Projeto orientado por:

Professora Doutora Helena Iglésias Pereira

Projeto co-orientado por:

Doutora Engenheira Elisabete M. Mourinho Arsénio Guterres de
Almeida

2015

Agradecimentos

A realização desta dissertação foi apenas possível através do suporte de professores, familiares, colegas e amigos. Deste modo, dado não ser possível agradecer a todos aqueles que me acompanharam neste percurso, serão destacados apenas aqueles que deram o seu contributo para o resultado final desta dissertação.

Quero expressar o meu agradecimento à Professora Helena Iglésias pelo seu constante apoio ao longo deste trabalho e pela sua disponibilidade, contribuindo de forma determinante para a minha evolução na área de estatística.

Agradeço à Sr^a Doutora Engenheira Elisabete Arsénio, do Laboratório Nacional de Engenharia Civil, pela disponibilidade demonstrada ao longo deste percurso, sempre que necessitei, e pela sua orientação.

À minha família, designadamente ao meu pai, à minha mãe, ao meu irmão e à minha irmã expresso a minha gratidão pela tolerância e por todo o apoio que me deram, nos momentos em que, por alguma razão, estava desmotivada, frustrada. A eles dedico este trabalho.

Agradeço aos meus colegas de mestrado, nomeadamente ao André Sintra, ao Diogo Costa e ao Vasco Alves, e também aos meus colegas do DEIO, designadamente à Cláudia Paradela, à Joana Tiago, ao Nuno Pinheiro e à Mariana Pereira pelos bons momentos de descontração que me proporcionaram ao longo desta fase, dando-me ânimo e força para a conclusão deste trabalho.

Agradeço também à Rute Veiga e à Maria João Sintra por me terem acompanhado nesta fase e pela vossa longa e forte amizade que me ajudou a superar alguns obstáculos que me foram surgindo ao longo desta caminhada. Ao Tiago André dirijo um agradecimento pelo seu constante incentivo para que eu nunca desistisse e pela sua paciência.

Índice

Agradecimentos	3
Índice de Tabelas.....	7
Índice de Figuras	9
Resumo	10
Abstract.....	11
Lista de Abreviaturas	12
Capítulo 1	13
Introdução	13
1.1 Apresentação e Enquadramento do tema	13
1.2 Objetivos	15
1.3 Metodologia	15
1.4 Estrutura do Inquérito	16
Capítulo 2	17
Mobilidade elétrica e Novas TIC.....	17
2.1. Perceção dos Atributos relativos à Bicicleta elétrica	17
2.2 Mobilidade versus Acessibilidade	20
2.2.1 Mobilidade Sustentável.....	20
2.2.2 Vantagens	21
2.2.3 Eficiência energética	21
2.3 Mobilidade Elétrica	21
2.3.1 Bicicleta elétrica.....	22
2.3.2 As diferentes classes de bicicletas elétricas existentes no mercado	22
2.3.3 Novas tecnologias inerentes à bicicleta elétrica	23
.....	25
2.4. A Mobilidade Elétrica em Portugal.....	26
2.4.1. Enquadramento.....	26
2.4.2. Objetivos	26
2.4.3. Produção e Eficiência Energética	26
2.5. Situação actual no uso da bicicleta elétrica em outros países	29
Capítulo 3	36
Enquadramento Teórico	36
3.1. Introdução.....	36
3.2 Modelo Linear Generalizado	37
3.2.1 Família Exponencial	38
3.2.2 Componentes do modelo linear generalizado.....	39

3.2.3	Estimação dos parâmetros do modelo	40
3.2.4	Inferência estatística sobre o MLG	43
3.2.5	Seleção do modelo	46
3.3	Modelo de Regressão Logística	48
3.3.1	Resposta binária.....	48
3.3.2	Interpretação dos parâmetros.....	50
3.3.3	Inferência sobre o modelo de regressão logística	52
3.3.4	Medidas de Qualidade	53
3.3.5	Análise de Resíduos	55
3.3.6	Capacidade preditiva do modelo	58
Capítulo 4	59
Análise Exploratória dos dados	59
4.1	A população e a amostra	59
4.1.1	Erro máximo absoluto	62
4.2	Tabelas de contingência	63
4.2.1	Análise de Independência/Associação	63
4.2.2	Tabela de Contingência 2 x 2	63
4.2.3	Tabela de Contingência r x k	65
4.2.4	Interpretação	67
4.3	Análise Global	67
4.3.1	Simplificar a mobilidade em duas rodas	67
4.3.2	Avaliação das preferências dos inquiridos de acordo com o regime de utilização.....	70
4.4	Análise local	71
4.4.1	Identificação da Escola	71
4.4.2	Género	72
4.4.3	Grupo socioeconómico	73
Capítulo 5	74
Modelação	74
5.1	Introdução.....	74
5.2	Opções de Mobilidade Sequenciais e Codificação das variáveis	75
5.3	Modelo base.....	76
5.3.1	Estratégias de Modelação	77
5.3.2	Qualidade de Ajustamento do modelo	77
5.3.3	Análise do modelo	78
5.4.	Modelo global	80
5.4.1	Escolha de covariáveis.....	80

5.5 Interações	83
5.5.1 Fatores de agrupamento	83
5.5.2 Método <i>Stepwise</i>	85
5.6 Análise de resíduos.....	87
5.6.1 Diagnóstico.....	87
5.6.2 Ajustamento do novo modelo	90
5.7 Predição do modelo.....	91
5.7.1 Erro de predição	91
5.7.2 Curva ROC.....	92
Capítulo 6	92
Interpretação dos resultados	92
6.1 Análise Exploratória	92
6.2 Modelo base obtido	94
6.1.1 Interpretação das covariáveis	94
6.1.2 Interpretação das Opções de Mobilidade Sequenciais	95
6.2 Modelo final obtido	97
6.2.1 Interpretação das covariáveis	98
6.3 Conclusões	102
Bibliografia.....	106
Apêndice 1	109
Apêndice 2	118
Anexo 1	119

Índice de Tabelas

Tabela 2.1: Evolução das patentes da bicicleta tradicional até à bicicleta elétrica [Mariño Pequini, Suzi (2000)]	20
Tabela 2.2: Categorias das bicicletas elétricas e sua descrição [Electric Bicycle Guide (2010-2013)]	23
Tabela 2.3: Inovações mais recentes na bicicleta elétrica, seu funcionamento e objetivo.....	24
Tabela 2.4: Características das bicicletas elétricas por Região/País [Navigant Research, 2014]	35
Tabela 3.1: Proporção de Sucessos/Insucessos versus Grupos de Covariáveis	49
Tabela 3.2: Critérios de discriminação.....	59
Tabela 4.1: Valores de $P(1 - P)$ e de $\sqrt{P(1 - P)}$	60
Tabela 4.3: Tabela de Contingência 2 x 2	64
Tabela 4.4: Níveis de Importância das barreiras relativamente à bicicleta elétrica ..	68
Tabela 4.5: Níveis de importância das barreiras relativamente à bicicleta convencional	69
Tabela 4.6: Tabela de Contingência (Escola versus Regime de Utilização).....	71
Tabela 4.7: Teste de Independência Qui-quadrado e Exato de Fisher	72
Tabela 5.1: Opções de Mobilidade.....	75
Tabela 5.2: Categorização das variáveis em estudo e a frequência de indivíduos por categoria	76
Tabela 5.3: Funções de Utilidade	77
Tabela 5.4: Critérios de Ajustamento	78
Tabela 5.5: Características Amostrais.....	78
Tabela 5.6: Medidas de Qualidade.....	79
Tabela 5.7: Sumário do modelo base I	79
Tabela 5.8: Sumário do modelo base II	80
Tabela 5.9: Modelos Simples ajustados para cada uma das covariáveis	81
Tabela 5.10: Modelo múltiplo com todas as covariáveis selecionadas (Modelo 1).....	82
Tabela 5.11: Modelo 2	83
Tabela 5.12: Modelo obtido com o método de Hosmer e Lemeshow (Modelo 3)	84
Tabela 5.13: Modelo 2 vs Modelo 3	85
Tabela 5.14: Modelo obtido com o método stepwise (Modelo 5)	86
Tabela 5.15: Modelo 2 vs Modelo 5	86
Tabela 5.16: Modelo obtido pelo método stepwise sem as 24 observações (Modelo 6)	90

Tabela 5.17: Modelo 5 vs Modelo 6	90
Tabela 5.18: Valores Ajustados (cutpoint = 0.5) vs Valores Observados	91
Tabela 6.1: Sumário do modelo base.....	94
Tabela 6.2: Perfis	95
Tabela 6.3: Perfis mais votados.....	96
Tabela 6.4: Perfis menos votados.....	96
Tabela 6.5: Sumário do modelo final.....	97
Tabela 6.6: Cruzamento da ausência das novas TIC com a Escola	100
Tabela 6.7: Cruzamento da presença das novas TIC com a Escola.....	100
Tabela 6.8: Cruzamento da ausência de ciclovias com a Escola	101
Tabela 6.9: Cruzamento da existência de 50% de ciclovias com a Escola.....	101
Tabela 6.10: Cruzamento da existência de 100% de ciclovias com a Escola.....	101

Índice de Figuras

Figura 1.1 Emissões de GEE nos transportes, em Portugal e na EU-28.....	14
Figura 2.1: Comparação dos custos energéticos entre diversas formas de transporte [Science of Cycling]	21
Figura 2.2: Descrição da roda de Copenhaga	25
Figura 2.3: Funcionamento da roda de Copenhaga	25
Figura 2.4 Vantagens de acordo com uma maior eficiência energética.....	27
Figura 2.5: Vendas na China e ao nível mundial, de bicicletas elétricas [Navigant Research, 2014]	31
Figura 4.1: Série Temporal relativo às populações anuais no município de Águeda [1981-2014].....	62
Figura 4.2: Série Temporal relativo às populações anuais no município de Águeda [2001-2013].....	62
Figura 4.3: Meio de transporte utilizado pelos estudantes.....	67
Figura 4.4: Atributo da bicicleta elétrica mais importante por parte dos estudantes	69
Figura 4.5: Regime de utilização da bicicleta elétrica de maior importância por parte dos estudantes	70
Figura 5.1: Resíduos Pearson padronizados vs Observações.....	87
Figura 5.2: Probabilidades Cruzadas vs Observações.....	88
Figura 5.3: Halfnormal dos resíduos de Pearson padronizados	88
Figura 5.4: Leverages vs Observações	89
Figura 5.5: Estatística D vs Observações.....	89
Figura 5.6: Curva ROC	92

Resumo

Atualmente a mobilidade elétrica, mundialmente, está em ascensão devido à necessidade, cada vez mais elevada, de se reduzir a emissão de gases de efeito de estufa. Portugal, sendo um dos países líder nas energias renováveis, possui os recursos necessários para a meta de eficiência energética, abrangendo o setor de transportes, para além dos sistemas de arrefecimento/aquecimento e produção de eletricidade. Relativamente ao setor de transportes e, em particular nas cidades, torna-se imperativo reduzir o uso dos veículos equipados com motor de combustão interna convencional promovendo, assim, a utilização de veículos elétricos (exemplo: bicicletas elétricas) como uma alternativa de transporte inovadora de baixo teor de carbono, mais benéfica para a saúde e com menores custos sociais ambientais. A presente tese tem como objetivo estudar o papel futuro da mobilidade elétrica de duas rodas nas deslocações de curta distância, designadamente nas deslocações casa/escola, e identificar quais os atributos das bicicletas elétricas considerados mais relevantes para os seus potenciais utilizadores. O presente estudo foi efetuado a partir da análise de um conjunto de dados recolhidos pelo LNEC em duas escolas secundárias localizadas no Município de Águeda (Escola Secundária Marques de Castilho e Adolfo Portela). A análise foi constituída por duas fases, sendo que na primeira foi feita uma análise exploratória ao conjunto de dados e na segunda foi utilizada uma metodologia estatística denominada de regressão logística, onde foram utilizados os *software SPSS version 22* e *R studio version 3.1*, respetivamente.

Na obtenção do modelo final de regressão logística, foram utilizadas diversas técnicas de seleção de covariáveis (atributos); nomeadamente o método de seleção de covariáveis *Hosmer and Lemeshow* e o método *stepwise*. Uma vez obtido o modelo foi necessário proceder a uma análise de resíduos, com o fim de se verificar o seu ajuste ao conjunto de dados, e avaliada a sua capacidade preditiva.

O modelo revelou 9 covariáveis associadas à variável resposta (*choice*), automóvel (*choice=0*) ou bicicleta elétrica (*choice=1*), que foram: diferença de custo entre a bicicleta e o automóvel, ausência ou existência de pista ciclável, a ausência ou presença das novas tecnologias de informação e comunicação, a diferença do tempo de viagem entre a bicicleta e o automóvel, a escola secundária, a experiência anterior em bicicleta, o nível socioeconómico, a interacção entre a escola secundária e as novas TIC e a interacção entre a escola secundária e a ausência ou existência de pista ciclável.

A análise efetuada sobre o modelo final de regressão logística permitiu identificar as covariáveis (atributos) associadas à escolha entre o automóvel e a bicicleta elétrica e saber quais aqueles que são determinantes para a escolha da bicicleta elétrica, tendo como alternativa o automóvel.

Palavras-chave: Mobilidade Elétrica, Veículos Elétricos, Bicicletas Elétricas, Modelo de Regressão Logística, Mobilidade Sustentável, Análise Estatística, Águeda (Portugal).

Abstract

Nowadays electric mobility, globally, is increasing due to the need of reducing the emission of greenhouse gases. Portugal, being a leader in renewable energy, owns the necessary resources to attain energetic efficiency, including the transport setor and also the cold/heating systems and electricity production. Regarding the transport setor, and particularly in cities, it is imperative to reduce the use of conventional internal combustion motor vehicles promoting, therefore, the use of electric vehicles (e.g. electric bikes) as an innovate low carbon transport option, healthier and with lower environmental social costs. Most of the studies showed that the promotion in using electric bikes (e-bikes) produces good results, fulfilling the goals related with health and environment. This thesis aims to study the future role of electric two-wheeled mobility when traveling short distance, particularly in home-school travel, and identify which attributes of electric bikes are considered more relevant. This study was based on the analysis of a collected set of data from two high schools located in the city of Águeda (Marques de Castilho High School and Adolfo Portela High School). After the explanatory data analysis, Logistic Regression Models were used to identify the covariates that may influence the choice of the electric bike. The software for this analysis were the *SPSS version 22* and *R studio version 3.1*, respectively.

Until the final logistic regression model was obtained, various techniques of selection of covariates (attributes) were used; particularly covariate selection *Hosmer and Lemeshow* method and the stepwise method. Once obtained the model, it is necessary a residual analysis, in order to check their fit to the data, and to evaluate their predictive ability.

The model revealed 9 covariates associated with the response variable (choice), car (choice = 0) or electric bicycle (choice = 1), which were: cost difference between bike and car, absence or existence of cycle path, the absence or presence of the new technologies of information and communication, the difference in travel time between bicycle and car, high school, previous experience in riding a bicycle, socio-economic level, interaction between high school and the new technologies of information and communication and interaction between high school and the absence or existence of cycle path.

Examining the final logistic regression model, it is possible to identify the covariates (attributes) which are determinants for the selection of the electric bicycle, when the car is the alternative mode available.

Keyword: Electric Mobility, Electric Vehicles, Electric Bicycles, Logistic Regression Model, Sustainable Mobility, Statistical analysis, City of Águeda (Portugal).

Lista de Abreviaturas

GEE - Gases de Efeito de Estufa

EU - *European Union*

LEZ - *Low Emission Zones*

ZER - Zona de Emissões Reduzidas

TIC - Tecnologias de Informação e Comunicação

LNEC - Laboratório Nacional de Engenharia Civil

ISEG - Instituto Superior de Economia e Gestão

BTT - Bicicleta Todo-o-Terreno

ONU - Organização das Nações Unidas

MOBI.E - *National Electric Mobility Network*

IOS - *Internet Operating System*

SMS - *Short Message Service*

ML - Modelo Linear

MLM - Modelo Linear Misto

MLG - Modelo Linear Generalizado

AIC - *Akaike Information Criterion*

Capítulo 1

Introdução

1.1 Apresentação e Enquadramento do tema

Ao longo das últimas décadas tem havido uma adesão crescente no uso individual de veículos movidos a combustível, fator que está na origem de problemas graves de poluição do ar e de congestionamento em muitas cidades europeias. A crescente dependência do setor dos transportes relativamente aos combustíveis fósseis, bem como a instabilidade ao nível de preços de mercado e a situação crítica de recursos naturais estarem em vias de extinção, devido ao seu elevado consumo, também representa uma adversidade para a mobilidade urbana na maioria das cidades europeias. A dependência no automóvel é de tal ordem que a sua utilização chega a ser desmesurada levando ao seu uso mesmo em deslocações de curta distância. Apesar do automóvel poder ser percecionado pela população como um sinal de desenvolvimento, este, por sua vez, também interage de forma negativa sobre alguns fatores ambientais e energéticos importantes designadamente, para maiores níveis de ruído, poluição do ar, consumo irracional de combustível (não aproveitando fontes de energia renovável) e o tempo adicional despendido no congestionamento rodoviário.

Por outro lado, o número de acidentes rodoviários nas zonas urbanas têm vindo a aumentar, sendo que os peões e os ciclistas, por serem os mais vulneráveis, são as principais vítimas. O crescimento notório do parque automóvel e do volume de tráfego nas cidades, do transporte rodoviário de mercadorias e do tráfego aéreo são os grandes causadores de emissão de dióxido de carbono, que por sua vez surge como uma ameaça aos objetivos de reduzir as emissões de gases poluentes com efeito de estufa (*GEE*) estabelecidos pela União Europeia (*UE*), no âmbito do Protocolo de Quioto. Em Portugal, em 2011 ocorreram 32541 acidentes de viação (com vítimas) em território continental, dos quais resultaram 42851 vítimas, evidenciando decréscimos de 8.1% e 9.4% respetivamente, em comparação com 2010. As vítimas mortais ascenderam a 891 (-4.9%), tendo-se registado 2265 feridos graves (-8.5%) e 43890 feridos ligeiros (-9.6%), face ao ano anterior. Em 2011, 24.7% das vítimas de acidentes de viação tinham idades compreendidas entre 35 e 49 anos, seguindo-se as vítimas de classe etária mais idosa: dos 50 aos 64 anos, com 16.5%, e 65 ou superior, com 13.9%. Assim como em anos anteriores, em 2011 os homens continuaram a representar a maior proporção de vítimas de acidentes de viação (57.3% do total).

Contudo, nas últimas duas décadas houve melhorias relativamente às normas dos combustíveis fósseis. Com efeito, a introdução de catalisadores e as melhorias na qualidade dos combustíveis, conduziram a uma redução significativa das emissões de óxidos de azoto provenientes de veículos de passageiros e de mercadorias. A eficiência ambiental dos automóveis também melhorou, dado que os carros, hoje em dia, consomem menos combustível e são menos poluentes do que os adquiridos há 15 anos. No entanto, muitas das melhorias ambientais mencionadas são anuladas pelo aumento exponencial do número de automóveis particulares em circulação e pelas maiores distâncias percorridas devido à ausência de associação entre o planeamento urbanístico e o planeamento de transportes. Atualmente há quase um automóvel para cada dois europeus, enquanto há cerca de 30 anos havia um para cada quatro.

Em Portugal, verificou-se uma subida acentuada das emissões de *GEE* originadas pelo setor dos transportes, até 2002, seguindo-se uma tendência decrescente, enquanto a média europeia apresentou uma subida, embora menos acentuada, até 2007. Embora este setor mostre sinais positivos na redução de emissões nacionais de *GEE*, continua a ser uma das principais fontes de emissões, representando 24.7% do total das emissões de *GEE* em Portugal em 2012, enquanto na EU-28 o mesmo setor representa 19.7% [APA, 2014 (65)].

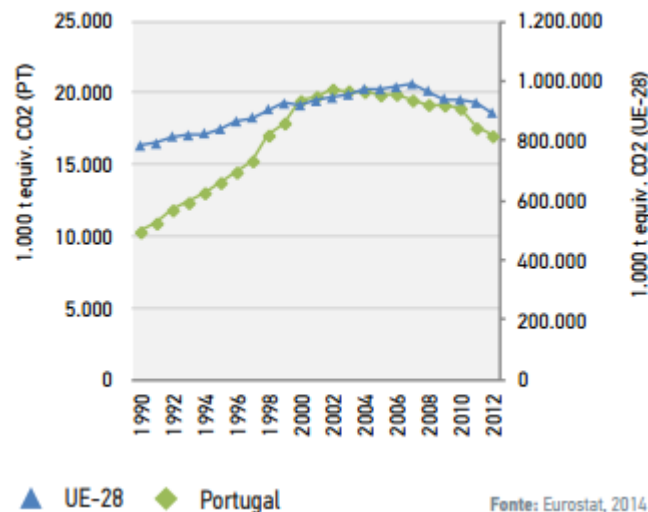


Figura 1.1 Emissões de GEE nos transportes, em Portugal e na EU-28

Em 1991 havia em Portugal, segundo o Eurostat, cerca de 175 veículos automóveis por 1000 habitantes. Em 2002 já haviam perto de 560, o que representa um crescimento astronómico de 220% em 11 anos, uma média de crescimento de 20% por ano. Por outro lado, de acordo com o INE no período de 2007 a 2011 a rede nacional de estradas cresceu 509 km na sua extensão, dos quais 288 km foram acrescidos no último ano, sendo este aumento correspondente a uma taxa de crescimento média anual de 1% no período dos últimos 5 anos. Por seu turno, a extensão da rede de auto-estradas expandiu-se a um ritmo ligeiramente superior (1.2%), passando de 2613 km em 2007, para 2737 km em 2011, apesar de não ter aumentado durante este último ano. Os distritos do Porto, Braga e de Lisboa são aquelas que apresentam os maiores índices de densidade da rede rodoviária, com valores próximos dos 3 km de rede rodoviária por cada 1000 km², ao contrário dos distritos de Beja e de Castelo Branco, cuja densidade da rede rodoviária não chega a 1 km (0.89 e 0.97 kms respetivamente). Em contrapartida, a concentração da rede rodoviária atendendo à população, foram os distritos de Beja e Portalegre que apresentaram os valores mais elevados, situando-se perto dos 6 km de rede rodoviária por cada 1000 habitantes, devido em grande parte à sua baixa densidade populacional. Por outro lado, os distritos com níveis de densidade populacional elevados (Lisboa, Porto, Aveiro e Braga) foram os distritos com menores índices de concentração de rede rodoviária por habitante sendo 0.36, 0.44, 0.73 e 0.98 kms por 1000 habitantes, respetivamente [Estatística de transportes - INE, 2011 (66)].

A circulação de veículos não seria tão caótica se a mesma fosse mais ordenada, podendo recorrer-se a regras de restrição ao uso do automóvel em determinadas áreas, como já acontece em vários países da Europa (*LEZ*). Por outro lado, o aumento da taxa de motorização pode induzir a expansão urbana e vice-versa, uma vez que as pessoas podem percorrer distâncias cada vez maiores para se deslocarem entre o local de residência, o local de trabalho, centros comerciais, a escola e os centros de lazer. Com vista a responder ao problema, têm sido criadas mais infraestruturas de transportes, nomeadamente mais auto-estradas, mas a teoria económica já comprovou que mais tráfego é por esta via gerado. Em suma, nos últimos 20 anos houve um aumento de 70% da rede de auto-estradas europeias, enquanto as linhas ferroviárias e fluviais diminuíram 9%.

Em Portugal também foram adoptadas medidas de restrição sobre o uso dos automóveis, tendo sido implementada na cidade de Lisboa a 1ª fase da Zona de Emissões Reduzidas (ZER) a 4 de Julho de 2011. Para além disso, honrando os compromissos assumidos em 2008 e pelo estado português, foi também aprovado o alargamento da área afeta à ZER, assim como o aumento da exigência ambiental e a redução das exceções, implementadas desde de 1 de abril de 2012. A 2ª fase tem por base o alargamento da área afeta à ZER, passando a existir duas zonas com restrições à circulação:

- Zona 1 - Residentes no Eixo Av. Liberdade/Baixa: restrição à circulação de veículos que não respeitem as normas de emissão EURO 2 (veículos ligeiros fabricados antes de Janeiro de 1996 e pesados antes de Outubro de 1996)

- Zona 2 - Residentes em Lisboa: restrição à circulação de veículos que não respeitem as normas de emissão EURO 1 (veículos fabricados antes de Janeiro de 1992)

O problema da qualidade do ar nas áreas urbanas tem vindo a causar cada vez maior atenção por parte das entidades comunitárias e do governo português, designadamente no eixo da Av. da Liberdade/Baixa pois é aquela que apresenta um índice mais elevado. Com a implementação das ZER é esperado um abate no parque automóvel, um incentivo crescente na utilização de veículos híbridos ou elétricos e uma transferência modal para os transportes colectivos, sendo o seu incumprimento incorrido numa multa.

Em Portugal, a cultura da utilização da bicicleta nas deslocações diárias em meio urbano é praticamente inexistente e, para além disso, os problemas ambientais e económicos relacionados com as fontes energéticas mais comuns (combustíveis fósseis, como o petróleo) são cada vez mais um motivo para que a poupança e a adesão às energias renováveis, mais sustentáveis, sejam cada vez mais uma realidade.

Com vista a combater o crescimento continuado da utilização do automóvel privado nos percursos de curta distância e promover o uso de opções de mobilidade urbana mais sustentáveis, o LNEC desenvolveu o projeto be4Schools no âmbito do planeamento da mobilidade elétrica e das novas tecnologias de informação e comunicação (TIC), o qual contou com o apoio do município de Águeda (Arsénio, 2013). Este estudo está em linha com o desenvolvimento de uma estratégia de mobilidade sustentável para o município de Águeda e constitui um estudo piloto do projeto de I&DT COST-TREND, financiado pela Fundação para a Ciência e Tecnologia, coordenado pelo LNEC, o qual envolveu uma parceria com o ISEG.

1.2 Objetivos

A presente dissertação, estando inserida no projeto do LNEC acima referido, pretende constituir um contributo para o desenvolvimento de uma estratégia de mobilidade sustentável no município de Águeda. Dar-se-á particular atenção à análise estatística e ao desenvolvimento de modelos econométricos tendo em vista avaliar o potencial das tecnologias de mobilidade elétrica emergentes, nomeadamente da bicicleta elétrica nas deslocações funcionais entre o local de residência e escola.

As questões que a presente investigação pretendeu dar resposta foram:

- Qual o papel futuro na mobilidade elétrica de duas rodas nas deslocações casa/escola?
- Que atributos/características da bicicleta elétrica são mais valorizadas pelas famílias?

1.3 Metodologia

Na presente dissertação considerou-se o inquérito à mobilidade que foi previamente desenvolvido pelo LNEC, de acordo com o estudo piloto acima descrito, o qual incluiu também um inquérito de preferências declaradas destinado ao potencial público utilizador da bicicleta elétrica, cobrindo estudantes dos 10º e 11º anos de duas escolas secundárias (escolas secundárias Adolfo Portela e Marques de Castilho) situadas no município de Águeda. Através do inquérito foram recolhidos vários tipos de dados e obtidas as opiniões e preferências dos estudantes relacionadas com as características dos veículos elétricos de duas rodas e de novos serviços de mobilidade previstos (bicicleta elétrica com inclusão de opções de novas tecnologias de informação e comunicação).

1.4 Estrutura do Inquérito

A análise estatística efetuada sobre o caso estudo tem como principal finalidade comparar aspectos associados à utilização das bicicletas elétrica e convencional, nomeadamente perceber qual a importância relativa do seu custo, modelo, os principais obstáculos ou barreiras à deslocação em cada tipo de veículo, etc. Este estudo também servirá de suporte à implementação da tecnologia elétrica no nosso país assim como à otimização dos custos envolvidos nesta opção modal de transporte. De acordo com os resultados do inquérito, os dados foram recolhidos em duas escolas secundárias, no Município de Águeda, tendo a amostra compreendido cerca de 350 inquéritos preenchidos (248 inquéritos foram considerados válidos no estudo). O inquérito foi organizado em três partes que em seguida se referem:

→ Parte I: Simplificação da mobilidade em duas rodas

Na primeira parte do questionário foram efetuadas perguntas direcionadas ao estudante, visando avaliar o uso preferencial do modo de transporte no percurso casa-escola, os atributos e preferências relativamente à bicicleta elétrica e as vantagens e desvantagens na utilização da mesma face à utilização da bicicleta convencional.

→ Parte II: Avaliação das preferências dos inquiridos de acordo com o regime de utilização da bicicleta elétrica

Numa 2ª parte, foi realizada uma análise, do ponto de vista económico, sob as preferências dos inquiridos, de acordo com o seu grupo socioeconómico, acerca do regime de utilização da bicicleta elétrica (tendo conhecimento de uma estimativa do preço actual de uma bicicleta elétrica). Foram também avaliadas as preferências de alguns modelos de bicicleta elétrica e de quatro opções de TIC passíveis de serem adicionadas.

→ Parte III: Análise das Opções de Mobilidade Sequenciais de Mobilidade Futuros

Por fim, foram apresentados 9 opções de mobilidade sequenciais para o uso da bicicleta elétrica especificamente em viagens de curta distância entre o local de residência e a escola. As 9 opções de mobilidade confrontam os inquiridos com 4 atributos de vários níveis cada um, sendo que a sua avaliação consiste em indicar se concordaria em substituir o veículo actual pela bicicleta elétrica na deslocação diária ou não. São considerados alguns pressupostos, nomeadamente:

- O estacionamento das bicicletas (disponível) é um atributo comum a todas as opções de mobilidade disponibilizados no inquérito.
- Estão à disposição bicicletas com um custo de aluguer mensal acessível (equivalente a um passe mensal).
- Conforme a distância entre o local de residência e a escola, assim varia a poupança obtida com a substituição do veículo ligeiro pela bicicleta elétrica.

Capítulo 2

Mobilidade elétrica e Novas TIC

2.1. Perceção dos Atributos relativos à Bicicleta elétrica

Desde o momento em que a bicicleta tradicional foi inventada que a sua indústria sempre se deparou, ao longo dos anos, com vários desafios, designadamente associados à evolução das preferências por parte dos utentes e por parte dos mercados relativos aos transportes e tecnologia. Todavia, esta dependência relativamente aos mercados mundiais não impediu que a procura deste modo de transporte aumentasse em determinados momentos da história nomeadamente durante as duas grandes guerras mundiais ou mesmo durante a crise petrolífera de 1973 que levaram à descida significativa dos níveis financeiros e económicos dos mercados.

Foi durante estas épocas que as populações procuraram rentabilizar os seus investimentos financeiros, tendo por base os recursos disponíveis e indispensáveis à melhoria da qualidade de vida. No caso do setor dos transportes, implicava a utilização do modo de transporte que fosse mais acessível na perspetiva económica, tendo por base os custos de aquisição do veículo e os relativos à sua operação e manutenção, sendo a bicicleta, assim, o veículo mais económico na perspetiva do utilizador.

Ao longo dos anos, com os avanços tecnológicos a surgirem cada vez mais sofisticados e eficientes, surgiram novas soluções com novos desempenhos. Esta situação levou a considerar a componente económica como uma dimensão importante para os utilizadores da bicicleta. Este veículo, ao longo da sua evolução, sempre foi influenciado pelo recurso de novas tecnologias, capazes de criar soluções eficazes que façam pôr à prova as aptidões e agilidades dos utentes como também toda a componente técnica do veículo de duas rodas.

A componente cultural também é igualmente tomada como um fator significativo na escolha deste meio de transporte, variando com o tipo de população conforme sejam naturais de países desenvolvidos ou de países em desenvolvimento.

Atualmente, com o aumento da população urbana em muitos países europeus e ainda em países em vias de desenvolvimento, existem outras consequências de carácter socioeconómico, dada a ausência de integração entre políticas de transportes e de desenvolvimento territorial e urbano, com um investimento centrado na construção de infraestruturas rodoviárias. Os espaços reservados para essa rede rodoviária nas cidades tornam impraticáveis a circulação segura de pessoas, dado que o seu projeto teve apenas em conta o tráfego de veículos movidos a motor de combustão interna. Para além disso, face ao elevado e instável preço dos combustíveis fósseis é também exigida uma reflexão sobre a possibilidade de se virem a utilizar outros modos de transporte de maior custo-eficácia ambiental, designadamente que conduzam à descarbonização do setor dos transportes. Devido à enorme dependência no uso de veículos ligeiros de passageiros, a redução do respetivo uso constitui uma prioridade para realizar progressos na direção da mobilidade sustentável. São nas deslocações de curta distância que o uso da bicicleta tem vindo a conquistar terreno, relativizando o congestionamento do tráfego. Do ponto de vista económico, é o meio de transporte mais rentável enquanto do ponto de vista ecológico, tem menores impactos ambientais (sem emissão de gases poluentes por exemplo, CO e CO₂), ruído de tráfego. Para além disso, estudos recentes indicam o uso regular da bicicleta como uma prática de exercício físico essencial [Marques da Silva, José Carlos, 2013 (2)].

Existem três tipos de bicicletas nomeadamente; a convencional, a assistida por motor de combustão interna e a assistida por motor elétrico. A grande vantagem das bicicletas elétricas ou motorizadas sobre a convencional é que estas possuem uma maior facilidade em

percorrer subidas sem qualquer esforço por parte do utilizador. Na tabela 2.1 apresenta-se a síntese relativamente à evolução histórica da bicicleta desde 1790 até à atualidade:

Ano	Evolução	Inventor
1790	Foi criada a 1ª patente da bicicleta denominada de celerífero. Ainda inviável como meio de transporte consistia numa trave de madeira prolongada por uma cabeça de animal e colocada sobre duas rodas, também de madeira, uma atrás da outra, com direção fixa. O arranque era dado com os pés firmes no chão, impulsionando, conjuntamente com a ajuda de uma 2ª pessoa que empurrasse o veículo pois este não tinha tração na roda como as atuais.	Conde Sivrac
1816	Criação da Drassiana: Com o fim de torná-la dirigível, foi acrescentado um guiador e molas para o assento. Surgiu então uma das primeiras bicicletas. Este modelo ainda apresentava alguns aspectos que requeriam uma solução técnica eficaz, nomeadamente a necessidade de criar um mecanismo de propulsão que não fosse através da força exercida pelos pés no solo.	Karl Drais von Sauerbronn
1817	Surgiu uma adaptação na Drassiana que consistiu numa inserção de um centro de gravidade muito mais baixo pois a barra que servia de trave e de assento estava muito baixa, possuindo um assento regulável visando um possível ajuste conforme a altura do condutor.	Georg von Reichenbach
1819	Foi criado um modelo Drassiano adaptado para as senhoras e que consistiu numa armação de madeira, coberta por ferro, com uma inclinação muito acentuada com a finalidade de se poder baixar, evitando o problema das suas saias compridas.	Dennis Johnson
1821	Surgiu a resolução de um mecanismo de propulsão alternativo ao uso da força dos pés no solo. Apesar de ainda sofrer um gasto significativo de energia, este mecanismo era constituído por uma manivela e uma roda dentada que impulsionava a roda dianteira.	Lewis Compertz
1838	Foi desenvolvido o mecanismo que ainda hoje é usado. A intervenção consistiu na junção de elementos no eixo da roda traseira onde, por meio de duas manivelas, esta era acionada através de dois pedais unidos à parte dianteira do quadro.	Kirkpark McMillan
1853	Foi criado um modelo com um par de manivelas à roda da frente e uma estrutura metálica em ambas as rodas, tornando-a numa bicicleta.	Philip Moritz Fisher
1861	Foram colocados dois pedais para facilitar a sua impulsão e mobilidade.	Pierre Michaux
1869	Foi substituída a roda de madeira pela roda de estrutura metálica. Surgem também os modelos de bicicletas com tração na roda traseira, proporcionando um avanço importante na técnica de tração com pedal.	Trefz

1869	Este avanço na substituição do pedal oscilante pelo pedal com manivela através de varas.	McMillan
1869	Foi criada a tração por meio de uma corrente contínua de transmissão.	Guilmet Meyer
1884	Na cidade de Coventry, no Reino Unido, foi produzida a bicicleta de segurança <i>Kangaroo</i> . Com vista a evitar as capotagens, este modelo possuía uma roda dianteira menor.	
1887	Foi criado o pneu a ar. Contudo, esta descoberta tinha um inconveniente pois quando havia um furo no pneu ou um outro fim em que o deixasse inutilizável o tempo de reparação era muito longo.	John Bloyd Dunlop
1890	Surgiram muitos pedidos de registros de patentes para bicicletas elétricas.	
1891	Por esta altura já existiam 5 mil ciclistas, só em França, devido à prova Paris-Brest decorrida no ano anterior. Passados 10 anos essa popularidade já atingia os 10 milhões.	
1895	Por esta altura surgiu a primeira bicicleta elétrica da história, no entanto era ainda muito pouco eficiente para o seu propósito dados os recursos e os conhecimentos para a época.	Ogden Bolton Jr
1903	Surgiu a criação do cubo com torpedo, com roda livre e freio a contra pedal. Este mecanismo para além de permitir que a bicicleta avance sem o uso dos pedais também funciona como um travão se se pedalar para trás.	
1939 1945	Por ocasião da 2ª guerra mundial, tornou-se necessário criar uma bicicleta dobrável, capaz de ser transportada às costas presa nas mochilas dos soldados.	
1940 1950	Na época do pós-guerra tornou-se cada vez mais importante adotar uma componente do <i>design</i> industrial, chamada “ <i>Styling</i> ”, que tinha como principal objetivo conferir aos produtos estruturas mais aerodinâmicas e sofisticadas. Por esta altura, não só o público-alvo já possuía um gosto mais direcionado para a velocidade como o marketing já começava a ganhar um papel importante na aceitação dos produtos no mercado.	
1960 1970	Surgiram muitas inovações com o lançamento da BMX. Em 1969, uma invenção que consistiu numa roda traseira de fricção foi reexaminada e desenvolvida. O seu dispositivo usava 4 motores fracionários, todos estes conectados através de uma série de engrenagens.	G.A.Wood Jr.
1980	Foi lançada a <i>Mountain Bike</i> (BTT).	
1990	Foi lançada a <i>Multi Sport Zipp Speed</i> . Por esta altura, também foram desenvolvidos os sensores de movimento e os controles de potência.	
1992	Foi criada uma bicicleta elétrica de nome <i>Zike</i> , que era	

	constituída por baterias de níquel-cádmio e por um motor de ímã permanente de 850g.	
1993	Só de 1993 a 2004 a sua produção aumentou cerca de 35%, o que fez com que, em 1995, a produção de bicicletas tradicionais diminuísse desde o pico de 107 milhões de unidades.	
2001	Os termos e-bike, bicicleta potência “Pedelec”, assistida a pedal e bicicleta assistida já eram usados com muita frequência para se referir às bicicletas elétricas.	
2007	Esperava-se que 10 a 20% de todos os veículos de duas rodas nas ruas de muitas das grandes cidades fossem de bicicletas elétricas.	

Tabela 2.1: Evolução das patentes da bicicleta tradicional até à bicicleta elétrica [Mariño Pequini, Suzi, 2000 (12)]

2.2 Mobilidade versus Acessibilidade

Com o aumento da utilização do transporte individual, a mobilidade tem aumentado significativamente contribuindo, desta forma, para uma maior flexibilidade de deslocação nas áreas urbanas. No entanto, hoje em dia, dado que uma percentagem significativa da população trabalha fora da sua área de residência e os seus respectivos horários de trabalho são coincidentes, são gerados problemas de congestionamento nas infraestruturas rodoviárias e ao nível de espaços de estacionamento. Assim, a promoção do uso da bicicleta é uma boa iniciativa dado que poderá contribuir para a melhoria da mobilidade urbana.

A bicicleta é também um modo de transporte que contribui para a inclusão social e para a melhoria da acessibilidade da população (exemplo: acesso a equipamentos de lazer, compras, etc. sem necessidade de custos operacionais de transporte adicionais).

2.2.1 Mobilidade Sustentável

A mobilidade sustentável tem como principal finalidade priorizar os meios de transporte coletivos, a pé e de bicicleta sobre o automóvel e promover o reordenamento dos espaços, de forma a reduzir as necessidades de deslocação e os seus custos sociais (internos e externos). A mobilidade sustentável pressupõe também a acessibilidade universal e tem associada benefícios ambientais ao nível da redução de gases poluentes, com efeito de estufa e do ruído de tráfego emitido na locomoção de veículos motorizados; benefícios para a saúde pública, prevenção da sinistralidade rodoviária, uma maior eficiência energética na área dos transportes e uma redução de custos de mobilidade das famílias. Por outro lado, a mobilidade sustentável amplia o conceito de transporte para o de comunicação, através da utilização de novas tecnologias.

A Organização das Nações Unidas (ONU) elegeu a bicicleta como o transporte ecologicamente mais sustentável do planeta, devido ao seu baixo impacto sobre o meio ambiente, pela ausência de ruído e de emissões de gases poluentes e pela infraestrutura necessária à sua circulação e estacionamento. É também um meio de transporte de fácil utilização, podendo servir de alternativa para a maioria das pessoas, sem graves deficiências, desde a infância à velhice. O transporte de duas rodas pode também complementar as viagens em transporte público, permitindo alargar a sua quota de mercado, e as deslocações a pé.

2.2.2 Vantagens

Ecologia

Com a produção de um elevado número de bicicletas, em alternativa à produção do mesmo número de automóveis, é poupada uma grande quantidade de matéria-prima. Para além do seu tempo de vida ser, normalmente, mais longo, quando o veículo de duas rodas chega ao fim de vida útil a sua reciclagem produz um custo muito baixo. Com a redução no consumo dos combustíveis fósseis, a emissão de gases poluentes será significativamente mais baixa e consequentemente o efeito de estufa não será aumentado.

Economia

Apesar do custo de aquisição da bicicleta e de equipamentos adicionais como o cadeado, luvas e capacete não ser barato, o uso da mesma acaba por constituir um modo mais económico de transporte em zonas urbanas. Onde o seu custo difere significativamente do automóvel é no custo de manutenção, já que pode ser efetuada pelo próprio utilizador. Para além disso, a economia do combustível e do estacionamento também contribuem positivamente para a referida diferença de custos.

Saúde

Como já havia sido referido, o uso da bicicleta pode servir de terapia para determinadas doenças cardiovasculares, pulmonares ou problemas ligados ao excesso de peso. O uso regular deste meio de transporte produz benefícios para o sistema cardiorrespiratório, para além de ser uma atividade física que conduz à boa forma de um modo mais fácil e com baixo risco de lesões.

Tempo de viagem

Relativamente ao tempo de viagem ou de deslocação, a bicicleta é um meio de transporte que permite economizar o tempo, pois, ao contrário do automóvel, evita o forte trânsito nas vias de comunicação das grandes cidades e a procura de estacionamento.

2.2.3 Eficiência energética

O uso de uma bicicleta convencional exige ao condutor cerca de 22 Cal/km, constituindo a forma mais económica de locomoção. Uma caminhada de 1km consome cerca de 62 calorias e um automóvel médio a gasolina com apenas um ocupante gasta cerca de 1150 Cal/km. Admitindo que uma bicicleta com o condutor pesa cerca de 90kg e um automóvel, nas mesmas condições, pesa cerca de 1500kg, tem-se no primeiro caso um consumo de 0.24 Cal/kg e no segundo caso um consumo de 0.76 Cal/kg. Assim, é possível constatar que a bicicleta mostra ser um meio de transporte muito eficiente [Science of Cycling].

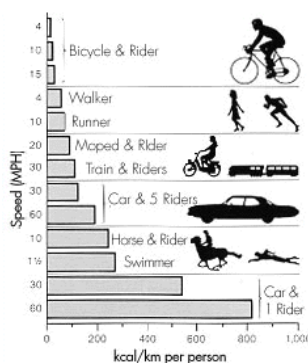


Figura 2.1: Comparação dos custos energéticos entre diversas formas de transporte [Science of Cycling]

2.3 Mobilidade Elétrica

Um outro objetivo por parte da UE e dos governos de Portugal passa pela adoção de novos modelos de mobilidade urbana, nomeadamente modos mais sustentáveis baseados na tecnologia elétrica e que reaproveitem a energia proveniente de fontes renováveis, podendo ser empregues no desenvolvimento da mobilidade sustentável nas cidades.

O uso da bicicleta elétrica poderá vir a ser a grande impulsionadora desse novo modelo de mobilidade, pois para além de depender do consumo às fontes renováveis, proporcionando uma melhor qualidade ambiental, vai também potenciar a aplicação de novas tecnologias no setor dos transportes. Com efeito, todas estas medidas vão ao encontro dos principais objetivos a serem superados sobre as questões da dependência energética e do combate às alterações climáticas, conforme consta no Livro Branco dos Transportes da Comissão Europeia (CE, 2011).

As principais finalidades com a implementação do Programa de Mobilidade elétrica, de acordo com a Resolução do Conselho de Ministros, são:

- Acelerar o processo de adoção de veículos elétricos e permitir a gradual conversão do parque automóvel;
- Incentivar a produção e desenvolvimento de produtos relacionados com a mobilidade elétrica, em Portugal, promovendo assim a evolução e o uso das novas tecnologias e inovações no país;
- Aproveitar a grande produção de energias renováveis existente no país.

2.3.1 Bicicleta elétrica

A bicicleta elétrica reúne as condições e vantagens de uma bicicleta convencional no que diz respeito aos potenciais benefícios para a saúde e ambiente, com a vantagem adicional de ser um veículo motorizado. Durante a sua deslocação, o ciclista tem um ganho no seu desempenho nos troços mais inclinados (subidas) com a ajuda de um motor elétrico, pelo que poderá percorrer maiores distâncias em termos relativos sem necessidade de esforço físico adicional. O motor, por sua vez, é alimentado por uma bateria que deverá ser recarregada regularmente.

Apenas com 250W é possível percorrer subidas, mesmo aquelas mais íngremes, dado que o esforço do ciclista é auxiliado pelo motor elétrico. Assim, a bicicleta elétrica permite ao utilizador uma deslocação mais rápida do que aquela efetuada por uma bicicleta convencional e menos poluente do que aquela efetuada por um automóvel. Para além disso, o utilizador deixa de se limitar a trajetos urbanos e pode atingir velocidades mais elevadas sem necessitar de um grande esforço físico.

2.3.2 As diferentes classes de bicicletas elétricas existentes no mercado

As bicicletas elétricas são classificadas de acordo com a potência que o seu motor elétrico pode suportar e com o seu sistema de controlo, isto é, quando e como a potência do seu motor é aplicado. Para além disso, as classificações das bicicletas elétricas são um processo algo complicado devido a razões legais sobre o que constitui um ciclomotor ou um motociclo. Estas razões variam muito conforme o país e a sua jurisdição local.

Em geral, as bicicletas elétricas podem classificar-se segundo duas situações; 1) aquela em que o motor da bicicleta assiste o utente usando o sistema de pedal auxiliar ou, 2) usando a potência acionada pela procura.

1. Com o sistema de pedal auxiliar, o motor elétrico é regulado pelo pedalar. Este sistema obriga a um maior esforço por parte do utilizador no uso da bicicleta. Estas bicicletas são as chamadas “*Pedelecs*”, possuem um sensor que deteta a velocidade dos pedais, a força destes ou ambos. Se o pedalar mantém o motor a trabalhar,

naturalmente se se deixar de pedalar ou se se usar o travão o motor é desativado. O termo *Pedelec* vem do inglês *pedal electric cycle* e o seu motor é relativamente de baixo consumo de energia. As *Pedelecs* são legalmente classificadas como bicicletas ao invés de motos de baixa potência ou ciclomotores.

2. Com a potência acionada pela procura, o motor é ativado por um regulador de pressão, geralmente montado num guiador, assim como na maioria das motos ou *scooters*. Estes veículos também denominados por *S-Pedelecs*, uma abreviação proveniente do alemão *Schnell-Pedelecs* ou do inglês *Speedy-Pedelecs*. Na Alemanha são classificados como ciclomotores ou motociclos. Estas bicicletas podem precisar de ser registadas e legalizadas, enquanto o piloto pode necessitar de uma licença de condução e de um capacete próprio. O seu motor elétrico é acoplado e operado manualmente, usando um regulador de pressão, que normalmente é no punho, semelhante ao sistema utilizado numa *scooter* ou numa moto. Normalmente possuem motores mais potentes do que as *Pedelecs*.

Categoria		Descrição
Bicicletas elétricas apenas com pedal auxiliar	<i>Pedelecs</i>	O seu motor auxilia só até a uma velocidade de cerca de 25km/h e a sua potência pode atingir os 250watts.
	<i>S-Pedelecs (Scooters)</i>	Possuem um motor com uma potência que pode ser superior a 250watts, podendo atingir uma velocidade superior a 45km/h antes de o motor deixar de auxiliar e sendo mais potentes que as <i>Pedelecs</i> .
Bicicletas elétricas com pedal auxiliar e com potência acionada pela procura		Algumas bicicletas elétricas combinam ambos os sensores de pedal auxiliar e de um acelerador. (Ex: Modelos <i>eZee Torq</i> e <i>Adventure 24+</i> de <i>BMEBIKES</i>)
Bicicletas elétricas apenas com potência acionada pela procura		Possuem um motor elétrico que é iniciado e operado manualmente através de um acelerador, usualmente situado no punho. As bicicletas elétricas com esta característica apresentam, quase sempre, um motor mais potente do que as <i>Pedelecs</i> .

Tabela 2.2: Categorias das bicicletas elétricas e sua descrição [*Electric Bicycle Guide*, 2010-2013 (60)]

2.3.3 Novas tecnologias inerentes à bicicleta elétrica

Com o surgimento da bicicleta elétrica em meados dos anos 90, foi uma questão de tempo até começarem a aperfeiçoar este meio de transporte conforme as necessidades da sociedade e o atual avanço tecnológico. Houve inovações criadas com o objetivo de reduzir o efeito de estufa, controlar o congestionamento, reaproveitar a energia elétrica libertada cada vez que

o utilizador pedala, aumentar a segurança do veículo e de criar uma rede de partilha de informação através das redes sociais.

De seguida é enunciada uma tabela, apresentando e descrevendo as inovações mais recentes do meio de transporte elétrico de duas rodas:

Nome	Funcionamento	Objetivo
Sensor especial que mede com precisão a força aplicada nos pedais	Pode ser visto no visor principal. O sinal deste sensor é enviado para um computador. Quanto maior for a força exercida pelos pedais, por parte do utilizador, maior será a potência fornecida ao motor elétrico.	Permite ao utilizador controlar a autonomia da bateria do veículo, impulsionando a força adequada.
Motor silencioso	Ao contrário dos veículos motorizados, os veículos movidos a baterias produzem um ruído sonoro nulo.	Reduz o ruído de tráfego no espaço público.
Níveis de Potência alternativos	Existem dois tipos: o modo eco (alcance máximo) que procura o mínimo possível e circular o máximo possível e o modo <i>boost</i> (potência máxima) que se torna útil em subidas de rampas ou qualquer plano inclinado que possa surgir na via.	Permite ao condutor, conforme os seus objetivos ou necessidades de viagem, escolher entre uma viagem económica ou rápida.
Chip eletrónico anti-roubo	Possui uma chave capaz de controlar tanto a bicicleta com o fecho de segurança da bateria deste. Com o uso desta chave, a bateria também pode ser programada para ser usada na bicicleta de um utilizador como na do seu parceiro.	Fornecer segurança ao veículo.
Bateria “invisível”	Situa-se por baixo da bagageira, sendo que a mesma não é visível do exterior também pelo motor estar localizado dentro do cubo da roda dianteira.	Contribui para um design muito mais sofisticado.
Roda que armazena energia e mede os níveis de poluição, o trânsito e o estado das ruas	À medida que o utilizador pedala, a energia libertada durante esse processo é direcionada para a roda reaproveitando-a para, por exemplo, usá-la onde as ruas têm uma inclinação acentuada.	Contribui para um aumento na eficiência energética e para um sistema de informação mais eficaz.
Sistema de comunicação entre um <i>Smartphone</i> e a bicicleta elétrica híbrida	Através do Bluetooth é possível bloquear ou desbloquear a sua bicicleta, ter acesso ao número de calorias gastas pelo utente, mudar as velocidades, definir o nível de energia assistida pelo motor e aceder aos níveis das condições atmosféricas, do tráfego, do relevo das ruas, dos níveis de CO e NO _x e de humidade.	Permite partilhar os dados com os amigos ou com qualquer outro interessado contribuindo para uma informação ambiental sólida e eficiente.

Tabela 2.3: Inovações mais recentes na bicicleta elétrica, seu funcionamento e objetivo

De seguida é ilustrado um exemplo das inovações mais recentes descritas na tabela 2.3:

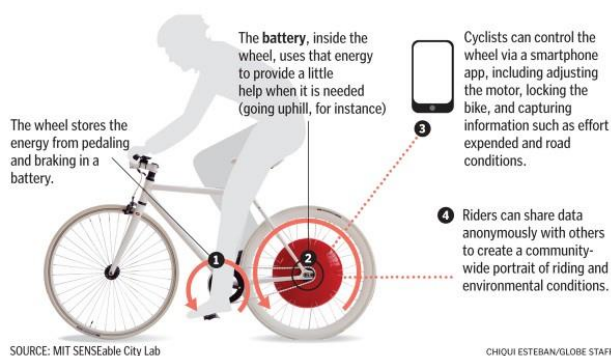


Figura 2.2: Descrição da roda de Copenhaga



Figura 2.3: Funcionamento da roda de Copenhaga

A roda de Copenhaga (*Copenhagen Wheel*) consistiu num projeto apresentado em 2009 pelo MIT SENSEable City Lab, tratando-se de uma roda especial, com novas funcionalidades, que é introduzida na parte da traseira da bicicleta. Cinco anos mais tarde, o projeto se desenvolveu e permitiu que uma bicicleta convencional pudesse se tornar elétrica, bastando substituir a roda traseira original pela roda de Copenhaga que integra um motor elétrico, baterias de íons de lítio e sensores de torque no eixo da roda.

Para além disso, através de um sistema sem fio, a jante está ligada ao *smartphone*, onde este deverá estar preso ao guiador, servindo como um painel digital que informa sobre a carga atual da bateria, a potência do motor e sobre a velocidade. O pedalar permite a roda funcionar, sendo que esta armazenará a energia adicional quanto mais forte for a pedalagem do utilizador. Essa energia adicional armazenada poderá ser aproveitada para percorrer ruas ascendentes. Através de uma aplicação do *smartphone*, os ciclistas podem ajustar o motor, bloquear a bicicleta e obter informação sobre o esforço físico despendido, as condições das ruas (ex: níveis de poluição atmosférica). Os ciclistas também podem partilhar informação, anonimamente, entre si, criando deste modo uma larga rede comunitária de ciclismo e de condições ambientais.

Relativamente às funcionalidades técnicas, a roda de Copenhaga tem motor elétrico de 350W na Europa e 250W na versão que será vendida nos Estados Unidos. A bateria opera a 48V e possui uma autonomia média de 50km/carga. A desvantagem mais relevante da roda de Copenhaga está relacionada com a sua compatibilidade, pois apenas pode ser instalada em bicicletas de aro 26, sem marchas ou com câmbio de 9 ou 10 velocidades.

A roda de Copenhaga possui um preço de pré-venda de cerca de 560€, sendo que o produto pertence à marca *Superpedestrian* [MotorDream, 2014 (61)].

2.4. A Mobilidade Elétrica em Portugal

2.4.1. Enquadramento

Induzidos pela necessidade de reduzir a crescente dependência relativamente ao consumo de combustível (petróleo) e pelo seu consequente impacto ambiental, Portugal procura combater estas adversidades apostando em novos modelos de eficiência energética para a mobilidade. A mobilidade elétrica surge, em 2009, como uma rede integrada de postos de carregamento de veículos elétricos existentes em cerca de 25 cidades em território nacional, explorada através de uma plataforma tecnológica (MOBI.E) que permite o abastecimento de qualquer veículo elétrico, mediante o uso de cartão de carregamento apropriado [MOBI.E, 2012 (8)]. Uma vez que Portugal é líder no mercado das energias renováveis, sendo que cerca de metade da sua eletricidade é proveniente de fontes renováveis, os veículos elétricos surgem como uma oportunidade de reaproveitamento de toda a produção dessa energia. Durante a noite, o consumo de energia acaba por ser mais reduzido, contribuindo para um melhor planeamento e uso otimizado de energia produzida a partir das fontes renováveis quando o carregamento é feito durante esse período. Através destas iniciativas e planeamentos foi desenvolvido um outro fator igualmente importante na mobilidade elétrica, designadamente a eficiência energética.

2.4.2. Objetivos

O principal objetivo da implementação de novos modelos de mobilidade urbana de maior viabilidade e eficiência energética é combater algumas adversidades relacionadas com a mobilidade, trabalhando no sentido de incutir cada vez mais o uso de veículos elétricos, usufruindo das energias renováveis, para uma melhor qualidade de vida nas grandes cidades. O combate às graves adversidades tais como; a crescente dependência dos combustíveis fósseis, a crescente emissão de gases poluentes e risco de efeito de estufa torna-se cada vez mais urgente para então poder construir uma política energeticamente mais sustentável. Em paralelo, toda esta iniciativa visa igualmente estimular a criação de novos modelos de negócio associados ao *cluster* da mobilidade, de potenciar o desenvolvimento tecnológico dos sistemas de redes de comunicação, das infraestruturas energéticas e de todos os componentes dos veículos nomeadamente, as baterias, as *powertrains* e materiais mais leves e recicláveis.

2.4.3. Produção e Eficiência Energética

Portugal é detentor de um dos maiores parques eólicos da Europa, de uma das maiores centrais fotovoltaicas do mundo e faz parte de um dos 4 países da União Europeia que promoveu a Diretiva 2002/91/CE com vista a obter Certificados Energéticos nos seguintes casos:

- Para adquirir licença de utilização em edifícios novos;
- Numa eventual reabilitação importante de edifícios existentes (custo > 25% do valor do edifício sem terreno);
- Aquando da locação ou venda de edifícios de habitação e de serviços existentes (validade do certificado: máx. de 10 anos);
- Regularmente (6 anos) para todos os edifícios públicos (de serviços) com mais de 1000 m².

Possui também em vigor um plano nacional para a eficiência energética e um plano nacional de barragens.

- O parque eólico do Alto Minho, com 120 turbinas, tem 240MW de capacidade instalada, com uma previsão anual de 530GWh, com capacidade para abastecer de electricidade cerca de 160.000 casas.
- A central de energia solar de Amareleja, em Moura, tem capacidade para abastecer as necessidades energéticas de 30.000 famílias, ou o equivalente a uma produção de

93.000MWh por ano. Esta central contribui, desta forma, para uma diminuição de emissão de CO₂ em mais de 89.000 toneladas por ano na atmosfera.

- A adesão à Diretiva 2002/91/CE revela que, desde janeiro de 2006, é obrigatória a instalação de painéis solares em edifícios novos. Esta medida contribui eficazmente na redução do consumo de energia no setor da habitação.
- Em Portugal, o plano nacional de eficiência energética prevê que haja uma redução no consumo de energia em cerca de 10% em 2015. Este planeamento irá contar com as adesões de várias entidades ligados a diversos setores de actividade com uma economia calculada em cerca de 12%.
- Quanto ao plano nacional de barragens, prevê-se que Portugal consiga ultrapassar os 7.000MW de capacidade hidroelétrica até 2020 e que os novos aproveitamentos hidroelétricos a implementar assegurem valores de potência instalada adicional na ordem dos 2.000MW. Este plano mostra que Portugal pretende usufruir da sua alta produção no setor hidroelétrico.

De seguida são apresentadas, num esquema, as vantagens relativamente à maior eficiência energética proporcionada pela utilização do meio de transporte elétrico:

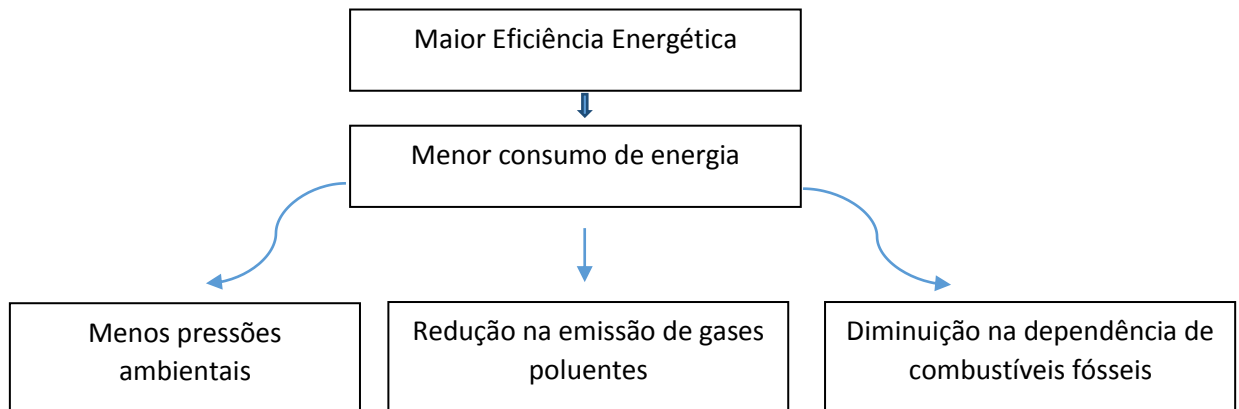


Figura 2.4 Vantagens de acordo com uma maior eficiência energética

A eficiência energética, em Portugal, foi potenciada através de uma rede de abastecimento espalhada por vários pontos do país, designadamente em 25 cidades [MOBI.E, 2012 (8)]. Esta rede consiste numa rede de carregamento inteligente acessível a todos os utilizadores que procura reabastecer os veículos elétricos de energia elétrica através do uso de cartão de carregamento apropriado. Para além disso, a rede MOBI.E também permite ao utente localizar e escolher postos de carregamento mais convenientes, planear os seus trajetos e controlar a autonomia da sua bateria. É possível aceder ao histórico de abastecimentos nomeadamente, a duração e o número de abastecimentos que foram efetuados até ao momento, os locais onde foram efetuados esses carregamentos e a quantidade de energia despendida. Com estas informações todas gravadas torna-se acessível ao utilizador poder planear de uma forma mais eficiente os seus gastos de energia, otimizando os custos de manutenção relacionados com os carregamentos.

- A rede MOBI.E dispõe de 1.300 postos de carregamento normal e 50 postos de carregamento rápido situados em espaços de acesso público em Portugal Continental.
- Os 1.300 pontos de carregamento estão situados em 25 cidades por toda a parte continental do país, nomeadamente Viana do Castelo, Aveiro, Bragança, Castelo Branco, Coimbra, Leiria, Lisboa, Cascais, Sintra, Setúbal, Évora, Beja, Faro, entre outros.

- Os 50 pontos de carregamento rápido estão distribuídos pelas vias de circulação entre estes concelhos, com via a uma maior eficiência no acesso entre estes, e também em locais estratégicos caso seja necessário um carregamento urgente.

Como já foi referido acima, existem dois tipos de carregamento; o normal (em corrente alternada) e o rápido (em corrente contínua), ambos com propósitos diferentes:

- O carregamento normal visa um carregamento diário, sendo mais indicado para os utilizadores que façam viagens de curta distância. O seu tempo de carregamento é no mínimo de 6 horas, podendo atingir por vezes as 8 horas de abastecimento.
- O carregamento rápido é usado habitualmente em situações de emergência ou quando o utente pretenda efetuar longos trajetos, de tal forma que a sua distância chegue a ser superior à autonomia do próprio veículo. A duração do seu abastecimento pode ser concluído entre 20 a 30 minutos.

Estes sistemas de carregamento só podem ser usufruídos e monitorizados por um cartão MOBI.E adequado, onde durante o seu carregamento é possível aceder ao Portal MOBI.E, através da sua área reservada ou através dos sistemas operativos para telemóveis, com as suas aplicações móveis disponíveis, como o *Android* ou *IOS*, e viabilizar a operação a decorrer. Através deste sistema de rede, é possível definir notificações de alerta, por SMS ou por correio eletrónico, quando o veículo fica carregado na totalidade. Uma vez atestado o veículo, todos os consumos e movimentos ficam registados no Portal MOBI.E para uma posterior consulta, é lançado um extrato de consumo com o valor do carregamento, incluindo o consumo de eletricidade e o montante relativo ao abastecimento.

Para além dos postos de abastecimento em espaços públicos, já referidos anteriormente, existem também os postos situados em espaços privados, embora cobertos igualmente pela rede MOBI.E são geridos por operadores privados. São exemplo disso locais como os parques de estacionamento, os centros comerciais, hotéis, aeroportos ou mesmo áreas de serviço. Existe também um sistema de carregamento doméstico, onde, por lei, é obrigatório que qualquer edifício novo construído ou reconstruído disponha de um ponto de carregamento ou de uma tomada elétrica para o carregamento de baterias de veículos elétricos. Para o caso dos edifícios que foram construídos antes do decreto-lei nº 170/2012 ter saído, os parques de estacionamento que tiverem tomadas elétricas com ligação exclusiva da fração poderão ser adaptadas para o carregamento elétrico. No entanto, torna-se imperativo verificar se a instalação elétrica é adequada ao serviço de carregamento, tendo que cumprir todos os requisitos técnicos e legais exigidos, pois os veículos elétricos apresentam exigências e especificações diferentes das dos eletrodomésticos podendo ser perigoso o uso indevido da instalação elétrica.

- Modos de Carregamento

A rede disponível em espaços públicos possibilita o carregamento de qualquer veículo nos pontos e lugares de abastecimento guardados para o efeito, contando com uma distribuição equitativa e uniforme em modo 3 com conectores tipo 2 (Mennekes) e tomadas modo ½ (cabo com control box) com conectores compatíveis (tomada industrial).

- Legislação

Em 2009 é criado o programa para a Mobilidade Elétrica em Portugal e o gabinete para a Mobilidade Elétrica (GAMEP), no Ministério da Economia e Inovação [MOBI.E, 2012 (8)]. Este programa apresentou os seguintes princípios fundamentais:

- 1- Entrar na perspetiva do utilizador, garantindo um acesso ao carregamento, independentemente do comercializador escolhido e assegurando as condições

- técnicas entre as diversas marcas e modelos de veículos, baterias e sistemas de carregamento;
- 2- O mercado relativo à mobilidade elétrica deve assegurar e apresentar condições que, de alguma forma, atraiam inúmeras empresas para a entrada do mercado, desenvolvendo desta forma a livre concorrência;
 - 3- O uso de energias renováveis será, neste programa, indispensável beneficiando dos mecanismos de produção descentralizada em meios urbanos.

Foi então estabelecido um pacote legislativo específico bem definido, apesar de flexível, em Abril de 2010, tendo sido adicionados vários incentivos diretos e indiretos para os veículos elétricos, apesar destes incentivos terem sido revogados em 2011 mediante a aprovação assumida no Memorando de Entendimento com a Troika.

A legislação portuguesa cria algumas vantagens para a utilização dos veículos elétricos, nomeadamente a isenção do pagamento do imposto sobre o veículo (IV) como também do imposto único de circulação (IUC). No caso das empresas, a aquisição de veículos elétricos poderá ser realizada deduzindo no IRC. As despesas com veículos elétricos estão isentas da tributação autónoma que se aplica aos veículos de empresa, sendo que esta isenção não se aplica nem no caso de veículos híbridos nem no caso de motores de combustão [Portaria n.º 467/2010]. Para além disso, as despesas com veículos elétricos estão isentos da tributação autónoma que se aplica aos veículos da empresa, exceto nos casos de serem veículos híbridos ou de serem veículos de motor de combustão. Relativamente à taxa de depreciação de veículos elétricos, está previsto um aumento dessa taxa face à dos veículos com motores de combustão interna.

2.5. Situação actual no uso da bicicleta elétrica em outros países

A expansão da mobilidade elétrica por todo o mundo é essencial na concretização vários objetivos, em particular daqueles que constituem preocupações à escala global como sejam a redução das emissões de gases com efeito de estufa. Em muitos países já existem medidas de política ligadas à mobilidade elétrica, visando impulsionar e incentivar o uso do veículo elétrico de duas rodas nas áreas urbanas. O mercado global de bicicletas elétricas continua a expandir-se com o surgimento de novas tecnologias e com o aumento da acessibilidade e da disponibilidade de ofertas de produtos. Muitos países da Europa e da Ásia tem desenvolvido mercados cautelosos devido às fortes culturas urbanas que utilizam as bicicletas elétricas como meio de transporte, designadamente a forte dependência na utilização do automóvel (falta de infraestruturas adequadas à circulação das bicicletas elétricas) e a segurança rodoviária onde, em particular, na China existe uma elevada densidade populacional e a utilização do meio de transporte elétrico de duas rodas é elevada, levando ao aumento no número de acidentes envolvendo os mesmos. Inversamente, as bicicletas elétricas ainda representam um mercado emergente na América do Norte, uma vez que estas têm sido utilizadas maioritariamente para fins de lazer do que como meio de locomoção. Os produtos de oferta no mercado das bicicletas elétricas variam consoante a região de venda. De seguida é efetuada uma descrição síntese sobre o mercado actual das bicicletas elétricas e respetiva legislação em vários países da Europa e do Mundo.

Dinamarca

Há cerca de 40 anos atrás, na cidade de Copenhaga circulavam exclusivamente veículos ligeiros de passageiros e de mercadorias mas hoje em dia 41% da população usa a bicicleta em viagens para as escolas e para o trabalho, por toda a área metropolitana. Cerca de 55% da população residente usam a bicicleta todos os dias, sendo que todos eles usam mais do que 1000km de ciclovias na grande cidade nos seus trajetos. Foram avançadas diversas medidas no sentido de tornar o seu uso sustentável, nomeadamente o planeamento do tráfego de bicicletas e da linha/trajeto desejável para os ciclistas e pedestres.

Atualmente existem muitas ruas de um sentido, na cidade de Copenhaga, onde alguns troços são aproveitados para servir de parque de estacionamento para as bicicletas e estes por sua vez servem de separadores entre a ciclovias e a rua reservada à circulação do tráfego

automóvel. Desta forma, são evitados muitos acidentes rodoviários (tendo havido uma redução de cerca de 35% em acidentes casuais envolvendo ciclistas em ruas particulares), limitando o espaço para a circulação dos automóveis, e com isso a segurança dos veículos de duas rodas aumenta, para além disso os utilizadores já possuem locais próprios onde deixar a sua bicicleta. Contudo, existem muitas outras cidades dinamarquesas onde foi reclamado espaço das infraestruturas rodoviárias para uso da bicicleta, nomeadamente nas cidades de Frederiksberg e de Nørrebro. O objetivo não só passa por reclamar espaço para estacionamento de bicicletas, reduzindo assim o espaço para estacionamento de automóveis, como também reaproveitar esse espaço plantando árvores.

Na cidade de Copenhaga foi lançado um projeto com vista a assegurar o estacionamento seguro de quatro bicicletas de carga, equivalente ao espaço de estacionamento ocupado por apenas um automóvel. Para tal criou-se o chamado *The Copenhagen Cargo Bike 'Car'*, baseado num carro cor-de-rosa visando reclamar espaço de estacionamento para esse género de bicicletas. Visto o seu número ter aumentado significativamente e cerca de 6% da população residente na cidade ter na sua posse pelo menos uma bicicleta de carga, havendo cerca de 30 000 bicicletas de carga na cidade (*Copenhagen.com*). Cerca de 25% das famílias com dois a três crianças possui uma bicicleta de carga e 50% dos utilizadores usam-no para transportar crianças. Apenas 22% dos utilizadores usam este meio de transporte em alternativa ao automóvel e 24% usam-no como suplemento ao automóvel. Para além disso, a propriedade do automóvel é muito baixa na cidade de Copenhaga, sendo que apenas 29% das famílias possuem um automóvel. A implementação desta iniciativa provocou um resultado positivo pois as pessoas estão a usar o meio de transporte de duas rodas com o fim de o substituir pelo automóvel e em viagens que antes eram efectuados pelo mesmo.

Por outro lado, em países como a Dinamarca a época de inverno é muito tempestuosa, com muita neve e com temperaturas negativas muito baixas, sendo o uso da bicicleta impraticável excepto quando acompanhado do equipamento certo para o efeito. No entanto, a aquisição de todo o equipamento indispensável incorre num custo de cerca de 870€, onde a maioria não tem condições para satisfazer essa garantia. De acordo com o Eurostat não existem leis nem regulamentos sobre o salário mínimo na Dinamarca, porém estima-se que o salário bruto anual na Dinamarca seja cerca de 38883€, isto é, cerca de 3240€/mês bruto. Para além disso, a Dinamarca é um dos países da Europa com maior taxa de renda (cerca de 55.4%), sendo o salário líquido anual cerca de 17342€, ou seja, cerca de 1445€/mês líquido. Apesar disso, existem locais que reaproveitam o seu equipamento de ski para poderem circular a bicicleta nas estradas quando neva.

Existe também um regulamento sobre as bicicletas, quanto à necessidade de reconstrução de estradas ou se o volume elevado de neve que fica acumulado, no sentido de manter o tráfego sustentável, dando prioridade às bicicletas. Para o efeito, são usados limpa-neves prioritariamente nas ciclovias e nos passeios e só depois nas estradas, para além de não ser permitido o bloqueio das ciclovias. Numa situação onde uma ambulância ou um carro dos bombeiros necessita de ocupar um espaço na via, por pouco tempo, é imperativo a existência de um sinal para informar os diversos tipos de tráfego/utentes da infraestrutura para evitar a sua utilização. Nestes casos, o automóvel fica limitado a circular em apenas em uma via da infraestrutura enquanto a bicicleta possui a sua via desimpedida.

China

Constitui, desde há muitas décadas, um país onde se usa regularmente o meio de transporte de duas rodas não só para viagens de lazer mas principalmente nas suas deslocações para o trabalho, os centros comerciais, pendulares ou mesmo para questões sociais. Para além do objetivo central de reduzir a emissão dos gases poluentes, a China apresenta um problema adicional relativo ao número elevado de acidentes rodoviários, na sua maioria, envolvendo utilizadores de bicicletas elétricas, convencionais ou peões devido à sua enorme densidade populacional (142 hab/km²), onde este possui mais de 1 bilião de habitantes.

São líderes mundiais no fabrico das bicicletas elétricas, com 22.2 milhões de unidades produzidas em 2009. A sua produção está concentrada em 5 regiões principais; Tianjin, Zhejiang, Jiangsu, Shandong e Shanghai, tendo chegado a exportar cerca de 370 000 bicicletas elétricas em 2009. O índice de vendas, não só das bicicletas elétricas sem assistência de pedal como também das *scooters*, atingiu as 56 mil unidades em 1998, chegando aos 21 milhões em 2008 e atingindo uma frota de 120 milhões de bicicletas elétricas no início de 2010. Este aumento significativo de vendas teve como principal impulsionador a exclusão e restrição, por parte das entidades governamentais chinesas, das motas nos centros das cidades, evitando, deste modo, perturbações no tráfego e acidentes. Deste modo, os utilizadores passaram a substituir as bicicletas convencionais ou as motas pelas bicicletas elétricas como meio de transporte pendular, em alternativa ao automóvel.

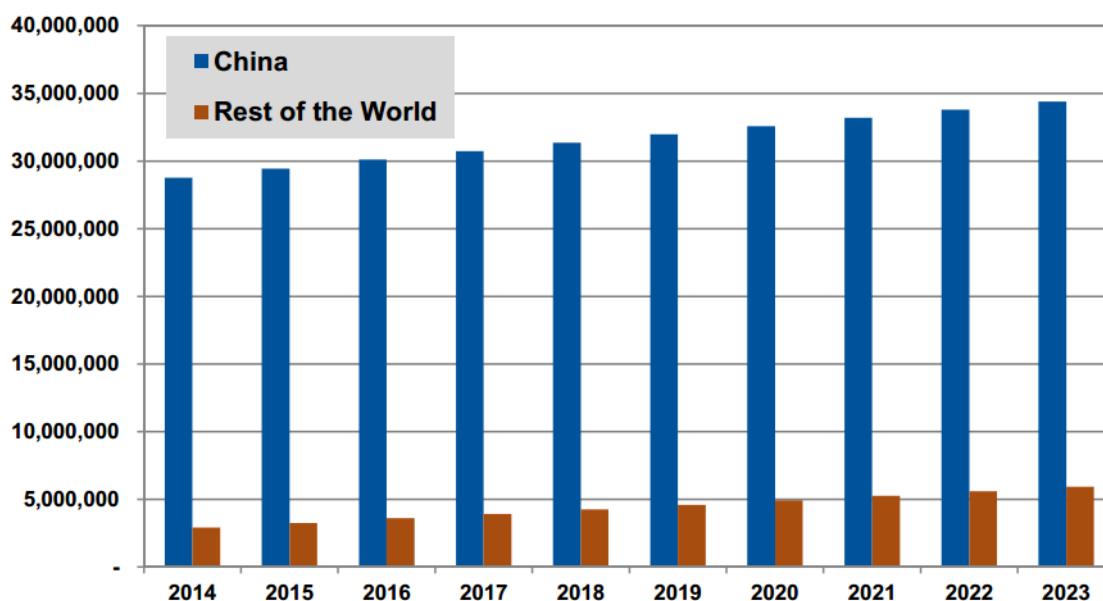


Figura 2.5: Vendas na China e ao nível mundial, de bicicletas elétricas [Navigant Research, 2014 (1)]

Contudo, por conta dessa elevada adesão, foram registados 2500 acidentes rodoviários envolvendo a morte de utilizadores do meio de transporte elétrico de duas rodas no ano de 2007. Pelo ano de 2009, 10 cidades também aderiram à banição ou restrição imposta no uso da bicicleta elétrica igual aquela que foi imposta sobre as motas, sendo as cidades envolvidas na causa Guangzhou, Shenzhen, Changsha, Foshan, Changzhou e Dongguang.

Relativamente à legislação, as bicicletas elétricas estão atualmente inseridas na mesma categoria das bicicletas convencionais, não sendo necessário uma licença de condução para as utilizar. Houve uma altura em que era obrigatório para o utilizador registar a sua bicicleta, visando ser recuperada em caso de roubo, no entanto esta medida foi recentemente abolida. Devido ao aumento de acidentes envolvendo utilizadores de bicicletas elétricas, causados principalmente por ciclistas inexperientes que circulam em sentido contrário nas vias, que atravessam o sinal vermelho, que não ligam os faróis de noite, etc., o governo chinês planeia legislar o uso das bicicletas elétricas de forma que as bicicletas que tenham uma tara de 20kg e ou mais e uma velocidade máxima de 30km/h ou mais apenas podem ser utilizadas com uma licença de mota, enquanto as bicicletas com uma tara inferior a 20kg e uma velocidade máxima inferior a 30km/h podem ser usadas sem licença. As bicicletas elétricas de velocidade inferior a 20km/h sem pedais são reconhecidas pela lei como um veículo de meio de transporte não mecânico, contribuindo deste modo numa promoção ao seu uso. De Agosto de 2002 a Janeiro de 2006, em algumas zonas de Pequim, foram banidas as bicicletas elétricas por motivos ambientais, de segurança e a respeito da imagem da cidade. A 4 de Janeiro de

2006 foram reautorizadas a circular pela cidade, no entanto existem ainda muitas cidades onde a bicicleta elétrica foi banida por lei.

De acordo com alguns estudos efectuados na cidade de Shanghai, visando analisar as principais características dos utilizadores de bicicleta elétricas, a maioria dos inquiridos (79%) deste modo de transporte elétrico de duas rodas pertencem ao grupo socioeconómico de médio e baixo rendimento. Ao longo da última década, o número de pessoas detentoras de bicicletas elétricas tem vindo a aumentar, sendo que em 2003 haviam cerca de 408 000 proprietários e em 2012 já constavam cerca de 2 797 000 de proprietários. As duas principais causas, que levaram à escolha das bicicletas elétricas como meio de transporte usual, são o baixo preço das mesmas e a maior garantia da pontualidade, dado os elevados níveis de congestionamento das cidades chinesas. A bicicleta elétrica é perfeitamente adaptável para utilizadores do género masculino ou feminino e de qualquer idade, dado não terem existido diferenças significativas no género ou na idade relativamente às características do uso do meio de transporte elétrico de duas rodas [Kang, An et al, 2013 (13)].

Relativamente aos principais motivos que levam os utilizadores a usar o meio de transporte elétrico de duas rodas, a grande maioria afirmou utilizar a bicicleta elétrica para viagens pendulares ou ir às compras (cerca de 79.2%). A proporção de viagens pendulares é mais baixa e a proporção de idas às compras é mais elevada, sendo que a bicicleta elétrica é o único meio de transporte dos inquiridos e é também o meio de transporte mais conveniente do que as restantes alternativas para se ir às compras. De acordo com a extensão das viagens pendulares, a bicicleta elétrica demonstra ser um meio de transporte perfeitamente adaptável em viagens pendulares de extensão até 15 km, tornando-se assim num forte concorrente aos autocarros de viagens pendulares de curta e média distância. Cerca de 12% dos utilizadores usam a bicicleta elétrica em substituição ao metro, em média, cerca de 2 vezes por semana.

A maioria dos utilizadores prefere usar a bicicleta elétrica devido à pontualidade, à poupança no tempo de viagem, ao seu baixo custo e à poupança no esforço. A sua preferência é relativamente à bicicleta elétrica em detrimento da bicicleta convencional e do autocarro pois a primeira é inadequada em viagens de longa distância e o segundo é muito caro, sempre lotado e impraticável durante as horas de ponta. Para além disso, cerca de 56% usavam a bicicleta convencional como meio de transporte comum antes de mudarem para a bicicleta elétrica, 33% usavam o autocarro e apenas 2% usavam o automóvel, sendo que este último fator pode indicar uma forte dependência e hábito por parte dos utilizadores do automóvel, nomeadamente devido à comodidade. Por outro lado, se o uso das bicicletas elétricas fosse proibido, cerca de 55% dos utilizadores iriam passar a utilizar o autocarro como meio de transporte, cerca de 31% passariam a usar a bicicleta convencional e apenas 7% mudariam para o automóvel, indicando que a presença da bicicleta elétrica contribui significativamente na redução das emissões de gases poluentes.

Num estudo efectuado na cidade de Shanghai, na China, verificou-se que as viagens de bicicleta elétrica, em média, são efectuadas num percurso de quase 10km e 30%, 37% e 33% dos inquiridos revelaram efectuar viagens pendulares com uma distância superior a 10km, entre 5 e 10km e inferior a 5km, respetivamente, conferindo ao veículo elétrico de duas rodas uma elevada capacidade de adaptação em percursos mais longos [Kang, An 2013 (13)].

Holanda

As bicicletas são um dos modos de transporte mais utilizados na Holanda, com uma quota de mercado de cerca de 27% das viagens urbanas e rurais, em território nacional. Nas cidades a percentagem é ainda mais elevada, atingindo os 38% na cidade de Amsterdão. Para além disso, existem cidades holandesas mais pequenas que excedem esse valor nomeadamente; a cidade de Zwolle com cerca de 46% e a cidade de Groningen com cerca de 59%, sendo provavelmente aquela com maior percentagem mundial. Todos os resultados positivos, no sentido de

promover o uso da bicicleta, foram conseguidos através da instalação de excelentes infraestruturas para a mobilidade de duas rodas, incluindo a partilha de espaço na mesma infraestruturas e ciclovias, acessos particulares adequados, cruzamentos seguros, criação de parques de estacionamento e da garantia de uma circulação mais rápida, curta e direta do que a dos automóveis.

Muitas infraestruturas rodoviárias possuem ciclovias em ambos os lados da via, sendo proibido a circulação de bicicletas na via principal na presença de ciclovias nessa mesma via. As ciclovias foram feitas com curvas ligeiras de forma a ser possível circular a uma velocidade considerável para distâncias longas, tendo também sinalização própria como sinais verticais, luminosos e tuneis. Contudo, nem todos os tuneis ou pontes podem ser atravessados por bicicletas, havendo autocarros ou comboios para o efeito que podem transportar bicicletas através do pagamento de uma taxa própria.

Quando existe um considerável espaço nas estradas, estas por sua vez são delineadas por uma ciclovia por meio de um ponto, cobertura, parque de estacionamento ou mesmo passeio para peões. Para além disso, podem haver ciclovias de dois sentidos e apenas um dos lados da estrada, reduzindo o número de vezes que o ciclista tem de atravessar a estrada principal para circular no sentido oposto. Existem também ciclovias marcadas na infraestruturas rodoviária principal, apenas separada por um tracejado branco e sendo distinguidas por terem uma cobertura vermelha. Alguns troços são constituídos apenas por ciclovias onde a circulação de bicicletas é prioritária (*Fietsstraat - bike street*), no entanto a circulação do automóvel ou de qualquer outro veículo a motor é permitida. Estes troços estão localizados principalmente em áreas residenciais onde existe menor tráfego rodoviário, são coloridos de vermelho como qualquer ciclovia usual e detém um limite máximo de velocidade 30km/h.

Na Holanda existe, por norma, um parque de estacionamento para bicicletas junto a qualquer estabelecimento comercial. Existem muitos parques de estacionamento de bicicletas por todo o país, alguns retendo milhares de bicicletas. Qualquer estação é vigiada regularmente, sendo removida qualquer bicicleta que não esteja devidamente estacionada e trancada. O cadeado é cortado e o utilizador incorre numa multa que pode atingir os 25€ para reaver o seu veículo. Esta regulamentação evita assim o estacionamento dos veículos de duas rodas nos passeios reservados ao público.

Existe também um sistema de partilha de bicicletas denominada de *OV-fiet (public transport bike)*, gerida pelas ferrovias holandesas (*Nederlandse Spoorwegen - Dutch Railways*). Usualmente estão localizadas em todas as estações de ferrovias, estações de metropolitano e paragens ferroviárias por todo o país. Em média, existem mais de 6000 bicicletas em 250 locais. Através de um cartão de utilizador, o ciclista incorre num custo anual fixo de 10€ e por cada reserva diária paga 3.15€. O programa de partilha, que está em vigor desde 2003, já arrecadou mais de 1 milhão de utilizadores no ano de 2011.

A Holanda possui uma frota de 18 milhões de bicicletas. As bicicletas elétricas já atingiram uma quota de mercado de 10% até 2009, dado que a sua venda quadruplicou de 40 000 unidades para 153 000 entre 2006 e 2009, e os modelos elétricos representam cerca de 25% das vendas totais de bicicletas consumidas nesse ano. Em 2010, uma em cada oito bicicletas vendidas no país é elétrica, renunciando o facto de que o custo das bicicletas elétricas é três vezes superior ao custo das bicicletas convencionais.

Em 2008, de acordo com um estudo de mercado, em média a distância percorrida por um utilizador de uma bicicleta convencional na Holanda é de 6.3 km, enquanto por um utilizador de uma bicicleta elétrica essa distância aumenta para 9.8 km. Concluiu-se também que a propriedade de bicicletas elétricas pertencia maioritariamente aos utilizadores com mais de 65 anos, sendo que a sua utilização é mais limitada.

Alemanha

A Alemanha tem melhorado significativamente as condições para a mobilidade em duas rodas, com o desenvolvimento de ciclovias urbanas e de trajetos de longa distância para a circulação de bicicletas. Atualmente possui uma rede extensa de locais devidamente assinalados e de ciclovias para distâncias mais longas. Nas principais vias das cidades já existem ciclovias devidamente separadas da estrada principal, no entanto grande parte dessas ciclovias são parte dos passeios pedestres já existentes pelo que é necessário um cuidado extra relativamente ao cruzamento com os peões e as possíveis curvas junto à estrada principal. Atualmente possui 200 ciclovias de longa distância que cobre a maior parte do país. A rede de ciclismo alemã é a rede da rota nacional de ciclismo da Alemanha. Atualmente 12 dessas rotas de ciclismo de longa distância, chamadas *D-Routes (Deutschland route)* atravessam diversos pontos do país, promovendo o turismo de bicicleta.

De acordo com uma organização industrial alemã, ZIV, as vendas das bicicletas elétricas representaram, no ano de 2013, cerca de 10% das vendas findadas em 2012. Foram vendidas cerca de 380 000 bicicletas elétricas, um aumento de 15% relativamente ao ano de 2011. Contudo, as vendas de bicicletas convencionais diminuíram cerca de 4% para 3.95 milhões de unidades no mesmo período, tendo sido referidas as más condições climáticas como o principal motivo na descida das vendas. Como resposta ao aumento de interesse nas bicicletas elétricas, o preço médio da bicicleta elétrica aumentou de 495€, em 2011, para 513€, em 2012.

Estados Unidos da América

Nos Estados Unidos da América, o meio de transporte elétrico de duas rodas tem sido promovido como uma alternativa “verde” de condução. Para além do principal objetivo dessa promoção ser a de reduzir as emissões de gases de efeito de estufa, a poluição do ar e o forte congestionamento, permite também contribuir para a redução significativa de obesidade existente nos americanos, onde quase 35% (mais de 78 milhões de adultos) da população adulta residente são obesos, entre 2011 e 2012.

Enquanto na Europa o motor da bicicleta elétrica só é iniciado com o pedal, nos Estados Unidos da América a maioria das bicicletas elétricas vêm equipadas com um regulador de pressão que permite ligar o motor sem ser necessário pedalar.

Um dos grandes entraves à promoção da utilização das bicicletas elétricas nos Estados Unidos da América é a falta de infraestruturas adequadas para o efeito. Dado que grande parte dos estados se encontram muito industrializados e com infraestruturas criadas para veículos mais rápidos como o automóvel ou o autocarro, a ideia de circularem bicicletas elétricas nestas condições pode ser potencialmente perigoso. Com efeito, existem atualmente algumas cidades onde a população residente se desloca para o seu trabalho de bicicleta elétrica e por este motivo foram construídas infraestruturas com o fim de os motivar ainda mais nesse sentido.

De acordo com a liga de ciclistas americanos, as viagens de bicicletas registaram um crescimento de cerca de 47% por todo o país, entre 2000 e 2011. Nas cidades de Chicago, Washington, Nova Iorque e Los Angeles estão a ser construídas ciclovias com o fim de tornar o seu uso pelas grandes cidades mais seguro. Assim como os carros elétricos, as bicicletas elétricas são caras, podendo o modelo mais básico custar cerca de 500\$ e um modelo mais robusto, com baterias de melhor qualidade, chegar a atingir um preço entre os 2000\$ e os 3500\$. No entanto, com o avanço rápido da tecnologia os preços podem baixar à medida que as baterias e os motores se vão tornando mais eficientes e as economias vão-se relativizando. Contudo, espera-se, de acordo com a base de pesquisa *Navigant*, um aumento na taxa composta de crescimento anual, entre 2014 e 2023, de 6.8% do mercado das bicicletas elétricas na América do Norte. Sendo que já ultrapassaram o Médio Oriente, tornando-se no terceiro maior mercado do mundo, e as suas vendas de bicicletas elétricas na região são previstas aumentar em quase 286 000 unidades em 2023.

De acordo com o Instituto de Transportes A&M do Texas, o congestionamento do tráfego incorre, para os Estados Unidos, num custo de cerca de 100 biliões de dólares anuais, tanto no tempo como no combustível. Na América do Norte, as bicicletas elétricas continuam a aumentar a sua quota de mercado global de bicicletas através da alta conscientização do consumidor e da diminuição dos preços de baterias de íões de lítio. As vendas estão a aumentar, em particular, nos Estados da Costa Oeste principalmente devido à maior consciência ambiental do consumidor, ao desenvolvimento de infraestruturas para as bicicletas e às condições climatéricas favoráveis.

Uma outra tendência está relacionada com a demografia, sendo que os utilizadores da bicicleta elétrica passam a estar representados por classes etárias mais jovens. Anteriormente, o meio de transporte de duas rodas era relegado às gerações mais idosas mas nos últimos anos tem havido um aumento de utilizadores com menos de 40 anos na América do Norte.

Nos Estados Unidos, de acordo com a lei federal de segurança do produto de consumo, as bicicletas elétricas de baixa velocidade são enquadradas na categoria dos veículos de duas ou três rodas movíveis exclusivamente através de pedais, com uma velocidade máxima de 20 mph (32km/h) quando impulsionada por um motor e com um motor elétrico que produza menos de 750 W. Desta forma, as leis para a circulação das bicicletas elétricas nas ruas públicas, passeios ou mesmo ciclovias estão sob a jurisdição de cada estado e variam conforme o mesmo. Para além disso, as bicicletas elétricas de baixa velocidade não classificadas como veículos a motor, dado possuírem as características já mencionadas quando o utilizador pesa cerca de 77kg. Desta forma, a bicicleta elétrica com estas características é sujeita à mesma regulamentação de produto de consumo dada à bicicleta convencional.

Região/País	Velocidade Máxima	Potência do motor eléctrico	Outros requisitos
Estados Unidos	Máximo 32km/h	Máximo 750W	Possuem pedais
Europa Ocidental	Máximo 25km/h	Máximo 250W	O motor é iniciado apenas com o auxílio do pedal
China	Máximo 20km/h	Sem limitação	Possui pedais; possui um peso máximo de 40kg

Tabela 2.4: Características das bicicletas elétricas por Região/País [Navigant Res., 2014 (1)]

Por outro lado, as bicicletas elétricas que excedam a potência e a velocidade máxima referidas são regulamentadas pelo Departamento Federal de Transporte e pela Administração de Segurança Rodoviária nas Estradas Nacionais como veículos motorizados e, como tal, estes necessitam de medidas de segurança adicionais. No entanto, tais necessidades não podem ser aplicadas às bicicletas elétricas montadas a partir de peças ou kits por um indivíduo. Nesses casos, as mesmas são regidas por leis estaduais e locais, dado que a lei federal não quer antecipar as jurisdições locais e estaduais de promulgarem as suas próprias leis que regem o funcionamento de tais veículos em vias públicas e estradas. A dita promulgação cita que: “Esta secção irá substituir qualquer lei estadual ou exigência no que diz respeito às bicicletas elétricas de baixa velocidade, na medida em que tal lei estadual ou exigência é mais rigorosa do que a lei federal ou exigências”.

Com o passar da legislação sobre a mudança e com a introdução de um programa de emissões de *cap-and-trade* (política de mercado com vista a proteger a saúde do ser humano e do ambiente, controlando elevadas quantidades de emissões de um grupo de fontes), o estado da Califórnia está a procurar alternativas para a redução de emissões no setor dos transportes. A adesão às bicicletas elétricas irá certamente ajudar nesse sentido, caso o seu uso for substituído pelo do automóvel. De acordo com um estudo acerca da experiência dos

utilizadores do meio de transporte elétrico de duas rodas, efetuado no estado da Califórnia, apenas 11% das viagens de bicicleta elétrica cobriam as viagens de veículos (autocarros e automóveis), enquanto cerca de 58% substituíam viagens a pé e 11% não substituíam [Popovich, Natalie et al, 2014 (15)].

Contudo, espera-se que o impacto sobre o uso do automóvel venha a aumentar através de um lançamento de um programa piloto de partilha de bicicletas elétricas na cidade de São Francisco. A maior melhoria relativa na saúde é o ganho na transição da inatividade para a atividade física moderada (Gojanovic et al., 2011), assim a substituição do automóvel pela bicicleta iria encorajar o benefício na saúde pública.

De acordo com os participantes nesse estudo, a característica mais importante na escolha da bicicleta elétrica sobre a convencional é a velocidade, no entanto esta não é mais veloz que o automóvel. Para além disso, acabam por ser mais caras, na sua aquisição e manutenção, e são geralmente mais pesadas do que as bicicletas convencionais. Enquanto a distância atingida por uma bicicleta convencional depende do esforço físico do utilizador, a distância percorrida por uma bicicleta elétrica depende da autonomia da sua bateria. Em comparação com o automóvel, para além de ser um transporte mais lento que este último, deixa o utilizador exposto às condições climáticas, ao tráfego, percorrem uma distância menor mas são mais baratos e produzem muito menos impactos ambientais. Dos 25 inquiridos que usavam o automóvel antes da bicicleta elétrica, 20 afirmaram conduzir muito menos o automóvel na presença da bicicleta elétrica. Todos os inquiridos que revelaram o custo monetário mensal incorrido em combustível antes e depois de terem adquirido uma bicicleta elétrica (15 de um total de 27) afirmaram que esse custo baixou após a aquisição da bicicleta elétrica, sendo que o custo, em média, passou de 91\$/mês para 60\$/mês.

Apesar disso, foram também apresentados alguns aspectos negativos (segurança, perigo de roubo e a pouca autonomia), relacionados com o uso da bicicleta elétrica, que necessitam de ser corrigidos antes desta se tornarem globalmente adotadas e, para além disso, existe a ideia de que os não utilizadores desconhecem o papel que a bicicleta elétrica pode possuir no seu dia-a-dia assim como o seu custo. Com efeito, no sentido de encorajar o uso da bicicleta elétrica, um programa educativo ou uma campanha elucidativa direcionada a grupos de indivíduos a quem a bicicleta elétrica lhe pudesse ser útil poderia ser um bom ponto de partida. Infraestruturas para bicicletas bem localizadas, instaladas e devidamente separadas do tráfego de automóveis iria melhorar o conforto e a segurança do seu uso, para além da sua legislação. Apesar do estudo ter sido efectuado sobre um número reduzido de participantes, sugere que o futuro da população de utilizadores de bicicletas elétricas pode vir a ser substancial e que os benefícios do meio ambiente irão certamente aumentar com o seu uso nos Estados Unidos.

Capítulo 3

Enquadramento Teórico

3.1. Introdução

Existem muitas situações onde a variável resposta não segue uma distribuição gaussiana (ex. escolha de um dado modo de transporte), por este motivo não é possível aplicar modelos clássicos como:

- o modelo linear (ML) quando as observações são independentes,
- e o modelo linear misto (MLM) quando as observações são não independentes.

Surgindo a necessidade de se estudar novos modelos no presente estudo, considera-se apenas a primeira condição, ou seja; a variável resposta segue uma distribuição não gaussiana e as observações são independentes.

Inicialmente, foi considerado o modelo multinomial *logit* pertencente ao grupo dos modelos de escolha discreta. Os princípios do modelo multinomial *logit* podem ser aplicados no caso estudo do presente trabalho, onde a variável resposta é binária representando as alternativas de manter o automóvel ($Y=0$) ou substituir este pela bicicleta elétrica ($Y=1$). Contudo, a sua aplicação insere-se, essencialmente, num conjunto de dados onde a sua variável resposta seja representativa de múltiplas alternativas, isto é, apresente mais do que duas alternativas como opção. Atualmente existem modelos direccionados especificamente para casos onde a variável resposta tenha apenas duas alternativas (resposta binária), nomeadamente os modelos de regressão logística.

Por este motivo, o modelo de regressão logística é a metodologia estatística mais adequada para a análise do caso estudo. Este modelo insere-se nos chamados modelos lineares generalizados (MLGs), que serão de seguida enunciados seguido do caso particular correspondente ao modelo de regressão logística.

3.2 Modelo Linear Generalizado

Inicialmente, a explicação de fenómenos aleatórios era efetuada através dos modelos lineares normais, mesmo quando esse fenómeno não respeitava os pressupostos da normalidade. Na tentativa de se corrigir essa condição, eram efetuadas algumas transformações adequadas à variável resposta. Aquela que ficou mais conhecida foi talvez a que foi desenvolvida por Box e Cox (Transformação Box-Cox, 1964),

$$z = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log(y) & \text{se } \lambda = 0 \end{cases}$$

sendo λ uma constante desconhecida e y a variável resposta positiva. A principal finalidade deste método, quando aplicada a um conjunto de valores observados, é produzir a simplicidade da estrutura do valor médio da variável resposta, da normalidade e da variância constante. Todavia, este método raramente resulta para apenas um único λ . Posteriormente foram efectuados avanços no conhecimento estatístico bem como no desenvolvimento computacional e com base nesse avanço surgiu a necessidade de se aplicarem novos modelos, como é o caso do modelo normal não linear. Mas foi por John Nelder e Robert Wedderburn (1972) que foi desenvolvida a proposta mais eficaz e inovadora, os modelos lineares generalizados (MLGs). Estes modelos são provenientes de uma síntese dos modelos lineares clássicos (ML) e de outros modelos. O objetivo principal é abrir um conjunto de alternativas para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como fornecer maior flexibilidade à relação entre o valor médio da variável resposta e o preditor linear. A mesma relação pode assumir qualquer forma monótona não linear, podendo dar-se a possibilidade de não ser a identidade.

Os MLGs foram formulados de modo a abranger a teoria de muitos outros modelos, nomeadamente os modelos de regressão linear, logística e de *Poisson*. Desta forma, este grupo de modelos permite englobar muitos modelos estatísticos, onde existe sempre uma combinação linear entre a variável resposta e as variáveis explicativas (covariáveis).

O MLG fornece muitas vantagens na sua aplicação, particularmente:

- Para além de abranger uma ampla classe de modelos de regressão, como já foi referido, permite modelar modelos de regressão para qualquer tipo de variável resposta, nomeadamente para variáveis contínuas, binárias, ordinais, multinomiais, contagens, para taxas e proporções, entre outros.
- A sua implementação nos diferentes modelos é simplificada pois pode ser aplicada na estimação, na inferência e no ajustamento do modelo.

Desta forma, os MLGs são uma extensão dos modelos lineares detentores dos modelos com a variável dependente (variável resposta) de distribuição não gaussiana. Só após o reconhecimento de que muitas das propriedades da distribuição gaussiana faziam parte das distribuições da família da exponencial, onde se encontram igualmente incluídas distribuições como a Bernoulli, a Binomial, a *Poisson*, a Exponencial, entre muitas outras, é que essa extensão foi alcançada.

A generalização dos MLGs foi obtida estendendo as hipóteses subjacentes aos MLs em duas direcções:

- a possibilidade da distribuição das variáveis resposta poder ser uma que não seja a normal,
- a relação entre a variável resposta e as variáveis explicativas poder ser não linear.

3.2.1 Família Exponencial

Designa-se por $Y = (Y_1, \dots, Y_n)^T$ o vector aleatório constituído por n variáveis aleatórias (v.a's) independentes. Diz-se que cada componente Y_i tem uma distribuição pertencente à família exponencial e escreve-se $Y_i \sim FE(a(\theta_i), \frac{\phi}{\omega_i})$ se a função densidade de probabilidade (f.d.p.) ou a função massa de probabilidade (f.m.p.) assumir a forma:

$$f(y_i|\theta_i, \omega_i, \phi) = \exp\left\{\frac{\omega_i}{\phi} [y_i\theta_i - a(\theta_i)] + c(y_i, \phi)\right\}$$

onde $a(\cdot)$ e $c(\cdot)$ são funções reais conhecidas, que determinam os membros da família de distribuições, θ_i e ϕ são os parâmetros escalares, e ω_i é uma constante conhecida que varia de observação para observação e à qual se designa de peso.

É pressuposto que a função $a(\cdot)$ é diferenciável e que o suporte da distribuição não depende de parâmetros desconhecidos.

Quanto aos parâmetros escalares, já mencionados, ϕ é um parâmetro geralmente conhecido, também representado por σ^2 , chamado de parâmetro de dispersão ou de escala, sendo constante ao longo das observações. Relativamente a θ_i , este representa a forma canónica do parâmetro de localização havendo a possibilidade de não ter uma família exponencial biparamétrica caso ambos os parâmetros sejam desconhecidos (Turkman, M.A. e Silva G.L. 2000). Para qualquer escolha do parâmetro de dispersão, ϕ , temos uma família exponencial.

Nas situações expostas acima, a família exponencial respeita as condições habituais de regularidade.

Valor Médio e Variância

Mostra-se que se Y_i tem uma distribuição pertencente à família da exponencial então tem-se que:

$$E(Y_i) = a'(\theta_i)$$

e

$$Var(Y_i) = \frac{\phi}{\omega_i} a''(\theta_i)$$

Desta forma, a variância de Y_i depende sempre do valor médio do mesmo, pois a função $a''(\theta_i)$ depende do parâmetro canônico de localização θ_i que, por sua vez, este também condiciona o valor médio. A função $a''(\theta_i)$ é a chamada função de variação ou de variância e será designada por $V(\mu_i)$, onde $V(\mu_i) = a''(\theta_i)$.

3.2.2 Componentes do modelo linear generalizado

Considere-se um vector de observações y , com n componentes, tomada como uma realização de uma variável aleatória Y (variável resposta) e onde as suas componentes são independentes com valor médio μ . Anexado a cada uma das componentes do vector de observações, y , tem-se um vector de $p \times 1$ covariáveis, X^T onde x_k representa a k -ésima covariável, e β um vector $p \times 1$ de parâmetros desconhecidos sendo, na maioria dos casos, $x_1 = 1$.

A parte sistemática (ou determinística) do modelo é uma especificação do vector μ em função de um reduzido número de parâmetros desconhecidos β_1, \dots, β_p . No caso dos modelos lineares, essa especificação é dada da seguinte forma:

$$\mu = \sum_{j=1}^p x_j \beta_j,$$

onde os β s são parâmetros cujos valores são usualmente desconhecidos e são estimados a partir do conjunto de dados. Se se tomar em consideração a indexação das observações ($i = 1, \dots, n$), nesse caso a parte sistemática é dada da seguinte forma:

$$E(Y_i) = \mu_i = \sum_{j=1}^p x_{ij} \beta_j; \quad i = 1, \dots, n$$

onde x_{ij} é o valor da j -ésima covariável da observação i . Na sua forma matricial (onde μ é uma matriz de $n \times 1$, X é $n \times p$ e β é $p \times 1$) fica:

$$\mu_i = x_i^T \beta; \quad i = 1, \dots, n$$

onde x_i^T é a matriz do modelo e β o vector dos parâmetros.

Em suma, o ML clássico pode ser definido em três partes:

- Componente aleatória: As componentes do v.a. Y são variáveis independentes e seguem uma distribuição normal com valor médio $E(Y_i) = \mu_i$ e variância σ^2 constante, isto é,

$$Y_i \sim N(\mu_i, \sigma^2) \quad i = 1, \dots, n$$

- Componente sistemática: as covariáveis x_i^T , para cada uma das observações, produzem um preditor linear η_i dado por:

$$\eta_i = x_i^T \beta; \quad i = 1, \dots, n$$

O preditor linear ao ser definido por $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ é subentendido que a relação entre o valor médio, μ_i , e o preditor linear, η_i , é a identidade.

- Ligação entre a componente aleatória e sistemática:

$$\eta_i = \mu_i;$$

Os MLGs podem ser estendidos de acordo com as hipóteses subjacentes ao ML em duas situações:

- A distribuição de cada uma das componentes do v.a. Y pode ser proveniente de uma qualquer distribuição pertencente à família exponencial, que não seja gaussiana,
- A função de ligação pode tomar um outro tipo de relação, que não a identidade, entre o preditor linear e o valor médio, ou seja,

$$g(\mu_i) = \eta_i$$

onde $g(\cdot)$ é uma função monótona, contínua e diferenciável designada por função de ligação.

De acordo com as três partes definidas para os MLs, os MLGs, sendo uma extensão dos MLs, são definidos da seguinte forma:

- Componente aleatória: As componentes do v.a. Y , dado o vector de covariáveis x_i , são variáveis condicionalmente independentes com distribuição proveniente da família exponencial,
- Componente sistemática: É produzido um preditor linear com a mesma estrutura que foi apresentada no caso do ML, nomeadamente:

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}; \quad i = 1, \dots, n$$

- Ligação entre a componente aleatória e a sistemática (função de ligação) é dada por:

$$g(\mu_i) = \eta_i$$

A função de ligação canónica é tal que

$$g(\mu_i) = \eta_i = \theta_i$$

Processos no ajustamento do modelo

Entre os processos de ajustamento disponíveis, destacam-se três; a estimação dos parâmetros do modelo, a seleção do modelo e a predição. É preciso ter em conta que não é assumida que uma análise consiste numa aplicação sucessiva destes três processos apenas uma vez, pois, na prática, existem retrocessos, falsos pressupostos que precisam de ser corrigidos, entre outros. O terceiro processo será desenvolvido na secção da análise de resíduos.

De seguida, irão ser enunciadas as principais propriedades de cada um dos três processos de ajustamento.

3.2.3 Estimação dos parâmetros do modelo

Em qualquer modelo, é imperativo estimar os parâmetros desconhecidos com o fim de se obter uma relação entre a variável resposta e as variáveis independentes. Neste caso, o parâmetro de interesse é $\boldsymbol{\beta}$ sendo a sua estimação baseada no método da máxima verosimilhança.

É assumido então que cada componente do v.a. Y tem uma distribuição pertencente à família exponencial, tomando a seguinte forma:

$$L = \prod_{i=1}^n L_i(\theta_i, \omega_i, \phi) = \prod_{i=1}^n f(y_i | \theta_i, \omega_i, \phi) = \exp \left(\sum_{i=1}^n \frac{\omega_i}{\phi} (y_i \theta_i - a(\theta_i)) + \sum_{i=1}^n c(y_i, \phi) \right)$$

Sendo o logaritmo da função verosimilhança $l = \log(L)$

$$l = \sum_{i=1}^n l_i(\theta_i, \omega_i, \phi) = \sum_{i=1}^n \frac{\omega_i}{\phi} (y_i \theta_i - a(\theta_i)) + \sum_{i=1}^n c(y_i, \phi)$$

A estimação do parâmetro β é calculada através das equações de máxima verosimilhança, de acordo com uma amostra aleatória com n observações e respeitando as condições apresentadas acima. A abordagem para a sua estimação consiste em determinar o máximo do logaritmo da função verosimilhança.

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j} = 0, \quad j = 1, \dots, p$$

Sendo l uma função composta dos β_j , $j = 1, \dots, p$, vem que:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$

Dado que $a'(\theta_i) = \mu_i$ e $a''(\theta_i) = \frac{Var(Y_i)\omega_i}{\phi}$, vem:

$$\begin{cases} \frac{\partial l_i}{\partial \theta_i} = \frac{\omega_i(y_i - \mu_i)}{\phi} \\ \frac{\partial \mu_i}{\partial \theta_i} = a''(\theta_i) = \frac{\omega_i Var(Y_i)}{\phi} \\ \frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \end{cases}$$

Donde

$$\frac{\partial l_i}{\partial \beta_j} = \frac{(y_i - \mu_i)x_{ij}}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}$$

Desta forma, as equações de máxima verosimilhança para β assumem a forma:

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \dots, p$$

Por fim, as equações de máxima verosimilhança para β também podem ser explicitadas da seguinte forma:

$$U_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \dots, p$$

onde U_j é a chamada função score. É um vector de dimensão p cujas componentes são $s_j(\boldsymbol{\beta}) = U_j = \frac{\partial l}{\partial \beta_j}$.

O conhecimento do parâmetro de dispersão é irrelevante para a estimação de $\boldsymbol{\beta}$, como já havia sido mencionado.

A matriz de informação de Fisher, também denominada matriz de variância-covariância da função score, é dada por:

$$I(\boldsymbol{\beta}) = E \left[-\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right]$$

É obtida considerando as segundas derivadas de l_i

$$-E \left[\frac{\partial^2 l_i}{\partial \beta_j \partial \beta_k} \right] = E \left[\frac{\partial l_i}{\partial \beta_j} \frac{\partial l_i}{\partial \beta_k} \right] = \frac{x_{ij}x_{ik}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

O (j,k)-ésimo elemento da matriz $I(\boldsymbol{\beta})$ é dado por:

$$-\sum_{i=1}^n E \left[\frac{\partial^2 l_i}{\partial \beta_j \partial \beta_k} \right] = \sum_{i=1}^n \frac{x_{ij}x_{ik}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

Na forma matricial,

$$I(\boldsymbol{\beta}) = X^T \tilde{W} X$$

com

$$\tilde{W} = \begin{bmatrix} \tilde{\omega}_1 & 0 & \dots & 0 \\ 0 & \tilde{\omega}_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \tilde{\omega}_n \end{bmatrix}$$

e

$$\tilde{\omega}_i = \frac{1}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

No entanto, as equações de máxima verosimilhança mencionadas não têm solução analítica, pelo que é necessário recorrer a métodos iterativos. Um dos métodos mais utilizados é o método dos mínimos quadrados ponderados. Para o caso em que a variável resposta segue uma distribuição gaussiana o método da máxima verosimilhança é exactamente o método dos mínimos quadrados ponderados.

Este método é baseado em cálculos computacionais que envolvem funções não lineares da resposta e onde o vector de pesos é recalculado em cada iteração (McCullagh, 1983). Aplicando o método iterativo dos mínimos quadrados ponderados obtêm-se o estimador de máxima verosimilhança de β no preditor linear, η (McCullagh & Nelder, 1989; e Azzalini, 1996). Este método é semelhante ao método de *Newton-Raphson* mas usa a quantidade de informação de Fisher em vez da matriz Hessiana. Sendo, por este motivo, também designado por scores de Fisher.

Com a estimação do parâmetro β , resta estimar o parâmetro de dispersão. Apesar deste último poder ser estimado através do método da máxima verosimilhança, existe um método mais elementar que proporciona geralmente resultados satisfatórios. Este método baseia-se na distribuição de amostragem da estatística de *Pearson* generalizada, para uma dimensão n suficientemente elevada.

Após o uso do processo iterativo para estimar o parâmetro β , pode-se calcular os estimadores de máxima verosimilhança para os parâmetros μ_i , $i = 1, \dots, n$, devido à propriedade de invariância dos estimadores de máxima verosimilhança.

$$\hat{\mu}_i = g^{-1}(\mathbf{x}_i^T \tilde{\beta})$$

Baseado na relação

$$Var(Y_i) = \frac{Var(\mu_i)\phi}{\omega_i}, \quad i = 1, \dots, n$$

O estimador ϕ é dado por:

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{\omega_i(Y_i - \hat{\mu}_i)^2}{Var(\hat{\mu}_i)}$$

A $\sum_{i=1}^n \frac{\omega_i(Y_i - \hat{\mu}_i)^2}{Var(\hat{\mu}_i)}$ dá-se o nome de estatística de *Pearson* generalizada.

3.2.4 Inferência estatística sobre o MLG

Uma vez obtidos os estimadores para o parâmetro de interesse, β , também designados de coeficientes, torna-se interessante inferir sobre os mesmos. Para o efeito é preciso tomar conhecimento da distribuição de amostragem de $\hat{\beta}$ todavia, em geral, nos modelos lineares generalizados não é possível obter as distribuições exactas, pelo que vão ser enunciadas as distribuições assintóticas que se verificam quando os modelos satisfazem certas condições de regularidade. É pretendido, nesta fase, avaliar a significância dos coeficientes do modelo, implicando a construção de testes de hipóteses e intervalos de confiança sobre os mesmos com vista a concluir se as covariáveis presentes no modelo são estatisticamente significativas. Por outras palavras, se as covariáveis inseridas no modelo estão estatisticamente associadas à variável dependente, Y .

Irá ser enunciado, primeiramente, a distribuição assintótica do estimador da máxima verosimilhança de β e as suas propriedades.

- O estimador de máxima verosimilhança de β é assintoticamente centrado sendo a matriz de variância-covariância aproximadamente igual a $I^{-1}(\beta)$.
- A distribuição assintótica de $\hat{\beta}$ é gaussiana p-variada com valor médio β e matriz de variância-covariância $I^{-1}(\beta)$ e escreve-se $\beta \sim N(\beta, I^{-1}(\beta))$.
- A estatística $(\hat{\beta} - \beta)^T I(\beta) (\hat{\beta} - \beta)$, conhecida como a estatística de Wald, segue uma distribuição assintótica qui-quadrado com p graus de liberdade (χ_p^2).
- A distribuição assintótica de $\hat{\beta}_j, j = 1, \dots, p$ é gaussiana com valor médio β_j e variância $I_{jj}^{-1}(\beta)$, elemento (j, j) de $I^{-1}(\beta)$.

Na prática como β é desconhecido e a matriz de informação de Fisher depende de β esta é desconhecida pelo que se substitui $I^{-1}(\beta)$ por $I^{-1}(\hat{\beta})$.

3.2.4.1 Testes de Hipóteses

A maior parte dos testes de hipótese sobre o vector β , podem ser formulados da seguinte forma:

$$H_0: C\beta = \xi \quad \text{vs} \quad H_1: C\beta \neq \xi$$

Onde C é uma matriz $q \times p$, com $q \leq p$ de característica completa q , e ξ é um vector de dimensão q previamente especificado.

Existem vários casos particulares a partir do teste acima mencionado, nomeadamente:

- Hipótese da nulidade de uma componente do vector parâmetro:

$$H_0: \beta_j = 0 \quad \text{vs} \quad H_1: \beta_j \neq 0,$$

para algum j , sendo neste caso $q = 1$, $C = (0, \dots, 0, 1, 0, \dots, 0)$ e ocupando o 1 a j -ésima posição e $\xi = 0$.

- Hipótese na nulidade de r componentes de β . Se tivermos, por exemplo:

$$H_1: (\beta_1 \dots \beta_r)^T = (0, \dots, 0)^T, \text{ então } q = r \text{ e } C = (I_r O_r \times (p - r)) \quad \xi = 0_r$$

onde I_r é a matriz identidade de dimensão r , $O_r \times (p - r)$ é uma matriz de zeros de dimensão $r \times (p - r)$ e O_r é o vector nulo de dimensão r .

Como já foi referido, qualquer uma das hipóteses anteriores retratam casos particulares do teste inicial. Cada uma das hipóteses corresponde a testar os coeficientes de submodelos a partir do modelo inicial considerado. A primeira hipótese consiste em testar apenas a significância do parâmetro β_j , associado à covariável x_j , num submodelo com todas as covariáveis do modelo original à excepção da covariável x_j . A segunda hipótese descreve o mesmo tipo de teste mas, desta vez, sobre um conjunto de r parâmetros de interesse, isto é, o novo submodelo não terá presente as r covariáveis definidas na hipótese nula do teste.

De seguida vão ser enunciadas as duas estatísticas mais usadas, entre três principais, para testar as hipóteses referidas anteriormente; a estatística de *Wald* e a estatística de Razão de Verosimilhanças.

Teste de Razão de Verosimilhanças

A estatística da razão de verosimilhanças é definida da seguinte forma:

$$\Lambda = -2\{l(\hat{\beta}_0) - l(\hat{\beta})\},$$

onde $\hat{\beta}_0$ e $\hat{\beta}$ são os estimadores de máxima verosimilhança de β , sob H_0 e sob $H_1 \cup H_0$ respetivamente.

A distribuição assintótica de Λ é, sob certas condições de regularidade e sob H_0 , qui-quadrado cujo número de graus de liberdade corresponde à diferença entre o número de parâmetros a estimar sob $H_1 \cup H_0$ e o número de parâmetros a estimar sob H_0 , isto é, r .

Valores elevados da estatística Λ conduzem à rejeição da hipótese nula, assim se o valor observado da estatística Λ for superior ao quantil de probabilidade $1 - \alpha$ da distribuição χ_r^2 , a hipótese H_0 é rejeitada ao nível de significância α .

Teste de Wald

A estatística de *Wald* é aquela que é, usualmente, aplicada para testar hipóteses nulas sobre cada um dos parâmetros individuais ou sobre um conjunto de componentes em simultâneo. No caso em que se testa sobre um conjunto de r componentes a sua fórmula é dada por:

$$W = (C\hat{\beta} - \xi)^T [CI^{-1}(\hat{\beta})C^T](C\hat{\beta} - \xi)$$

que, sob a hipótese nula, segue uma distribuição assintótica de um qui-quadrado com r graus de liberdade, χ_r^2 . Rejeita-se a hipótese nula, ao nível de significância α , se o valor observado da estatística de teste for superior ao quantil de probabilidade $1 - \alpha$ do χ_r^2 .

Por vezes pode interessar testar apenas para um único coeficiente do modelo, sendo que a sua fórmula é representada da seguinte maneira:

$$W = (\hat{\beta}_j - \beta_j)^T I_{jj}(\hat{\beta})(\hat{\beta}_j - \beta_j)$$

assim, sob H_0 , assume a forma $W = \frac{\hat{\beta}_j^2}{I_{jj}^{-1}(\hat{\beta})}$ e segue uma distribuição assintótica de um qui-quadrado com 1 grau de liberdade, χ_1^2 , uma vez que se está a testar apenas 1 coeficiente. A hipótese nula é então rejeitada ao nível de significância α se o valor observado da estatística de teste for superior ao quantil de probabilidade $1 - \alpha$ do χ_1^2 .

Sob o pressuposto da existência de normalidade no conjunto de dados (caso mais comum), a estatística de teste de *Wald* é dada por:

$$\sqrt{W} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{I_{jj}^{-1}(\hat{\beta})}}$$

onde, sob H_0 , a estatística \sqrt{W} segue uma distribuição *t*-student com $n - p$ graus de liberdade quando o parâmetro de dispersão é estimado, e uma distribuição gaussiana padrão ($N(0,1)$ -média nula e variância unitária) se o mesmo parâmetro for conhecido. Este resultado permite uma base na inferência exacta em amostras de qualquer dimensão.

Por outro lado, sob um pressuposto assintótico de segunda ordem tendo em conta o valor médio, a variância e covariância das observações, para amostras de dimensão elevada a estatística \sqrt{W} segue uma distribuição gaussiana padrão. Este resultado permite uma base na inferência assintótica em amostras de dimensão elevada.

3.2.4.2 Intervalos de Confiança

O intervalo de confiança para os parâmetros $\beta_j, j = 1, \dots, p$ ao nível de confiança $1 - \alpha$ podem ser obtidos com base na estatística de Wald, através da seguinte expressão:

$$\left(\hat{\beta}_j \pm z_{1-\alpha/2} \sqrt{I_{jj}^{-1}(\hat{\beta}_j)} \right)$$

onde $z_{1-\alpha/2}$ é o quantil de $(1 - \alpha/2)$ para a distribuição gaussiana padrão e $\sqrt{I_{jj}^{-1}(\beta_j)}$ é o desvio padrão do estimador de ordem j .

Para o vector β de dimensão p :

$$(\hat{\beta} - \beta)^T I(\beta) (\hat{\beta} - \beta) \leq \chi^2_{(1-\alpha, p)}$$

onde $\chi^2_{(1-\alpha, p)}$ é o quantil de $(1 - \alpha)$ do χ^2 com p graus de liberdade, dá-nos o elipsoide de confiança para β .

3.2.5 Seleção do modelo

Os modelos que são selecionados no ajustamento do conjunto de dados são usualmente escolhidos a partir de uma classe específica e para que o modelo seja representativo é preciso que essa classe seja amplamente importante para o tipo de dados no caso estudo. Uma característica importante nos modelos lineares generalizados é a assumpção de independência entre as observações (ou ausência de correlação). Esta assumpção de independência é uma característica proveniente dos modelos lineares de análise de regressão clássica e é transportada para os modelos lineares generalizados sem qualquer alteração.

A escolha de uma escala para a análise é um aspecto muito relevante para a seleção do modelo. Uma decisão comum, nestes casos, está entre tomar uma análise sobre Y , a escala original, ou sobre $\log(Y)$. Com a introdução dos modelos lineares generalizados, os problemas derivados da escala são significativamente reduzidos. A existência de normalidade e a variância constante já não são imperativas, apesar da relação dependente entre a variância e o valor médio ter de ser conhecido.

Outro factor muito importante no processo de seleção do modelo é a escolha das covariáveis a serem introduzidas na parte sistemática do modelo. Para o efeito, existem estratégias de escolha que consistem num balanço entre melhorar o ajustamento dos valores observados ao adicionar uma covariável ao modelo e o indesejado aumento da complexidade com a introdução dessa nova covariável. É imperativo ter em conta “*à priori*” a ideia de que existem várias alternativas para a escolha do modelo ótimo, visto ser pouco provável que o conjunto de dados indique um possível candidato à altura entre um elevado número de modelos significativos.

Numa abordagem inicial ao problema de modelação, é preciso encontrar um ou mais conjuntos parcimoniosos de covariáveis correspondente a uma matriz de dimensão $n \times p$, onde n é o número de observações e p o número de covariáveis do modelo. O termo parcimonioso, na área da estatística, implica que as covariáveis, que não produzem qualquer efeito na variável resposta, devem ser excluídas do preditor linear. Para que o modelo seja representativo, é necessário que não lhe sejam introduzidas interacções sem que produzam efeitos relevantes nem algum termo de grau superior sem o seu termo associado de grau inferior.

Seleção de Covariáveis

A seleção de um conjunto de covariáveis uteis para o modelo, a partir de um conjunto razoavelmente grande de possíveis covariáveis, com vista a formar uma estrutura parcimoniosa não é um processo trivial.

Método Stepwise

Existe um método que ignora qualquer relação existente entre as covariáveis e procura o melhor conjunto, de dimensão r , de covariáveis. Se $k \leq 12$, sendo k o número total de covariáveis disponíveis, o melhor conjunto de covariáveis, para cada componente r de 1 até

$k - 1$, pode ser encontrado passo a passo. Para $k \geq 35$, existem três métodos de pesquisa admissíveis para o problema de seleção. Os métodos de aproximação para gerar um único conjunto de covariáveis “ótimo” são os seguintes:

- *Forward selection*: Consiste num método onde em cada fase vão sendo inseridas as melhores covariáveis que satisfazem o critério definido até não restar mais nenhuma covariável nessas condições.
- *Backward elimination*: Método que começa com o modelo saturado, isto é, o modelo com o número máximo de covariáveis disponíveis e vai eliminando as piores covariáveis até todas as covariáveis que restem satisfizerem o critério.
- *Stepwise regression*: Método que engloba ambos os métodos anteriores, começando por *backward elimination*, seguido de *forward selection* e assim sucessivamente até se obter o melhor modelo de acordo com o critério definido.

Estatísticas para a seleção do modelo

Por outro lado, existem métodos de seleção de modelos com base na comparação de dois modelos distintos no número de covariáveis. Quando se pretende comparar dois modelos e decidir qual deles deve ser rejeitado, estes podem estar relacionados de duas maneiras:

- Ambos os modelos estão encaixados
- Ambos os modelos não estão encaixados

Para o primeiro caso, a função de desvio (*Deviance*), enunciada de seguida, pode ser aplicada.

Função de desvio

Na prática o modelo nulo (modelo com apenas uma variável) é usualmente demasiado simples e o modelo saturado (modelo com tantas covariáveis quantas as observações) demasiado complexo e pouco representativo uma vez que não resume o conjunto de dados e contém certamente muita informação redundante. No entanto, o modelo saturado proporciona um modelo base para medir a discrepância existente relativamente a um modelo intermédio com p parâmetros.

Sejam $\hat{\beta}_S$ e $\hat{\beta}_C$ os estimadores da máxima verosimilhança de β para o modelo saturado e para o modelo corrente, respetivamente. A estatística de teste de verosimilhanças já descrita anteriormente, é definida por:

$$\Lambda = 2\{l(\hat{\beta}_S) - l(\hat{\beta}_C)\} = -2\{l(\hat{\beta}_C) - l(\hat{\beta}_S)\}$$

podendo ser escrita também na forma:

$$= -2\{l(\hat{\beta}_C) - l(\hat{\beta}_S)\} = -2 \sum_{i=1}^n \frac{\omega_i}{\phi} \{[y_i \hat{\theta}_{C_i} - a(\hat{\theta}_{C_i})] - [y_i \hat{\theta}_{S_i} - a(\hat{\theta}_{S_i})]\} = \frac{D(\mathbf{y}, \hat{\mu})}{\phi}$$

onde $\hat{\theta}_{C_i}$ e $\hat{\theta}_{S_i}$ são os estimadores de máxima verosimilhança de θ_i para os modelos corrente e saturado, respetivamente e onde $D(\mathbf{y}, \hat{\mu})$ é a função desvio para o modelo corrente, sendo uma função apenas relativa ao conjunto de dados. Esta medida pode ainda ser descrita como a diferença entre os logaritmos das verosimilhanças observada e ajustada para a i -ésima observação. A razão entre a função desvio para o modelo corrente e o parâmetro de dispersão, $\frac{D(\mathbf{y}, \hat{\mu})}{\phi}$, é denominada de desvio reduzido.

$$D(\mathbf{y}, \hat{\mu}) = \sum_{i=1}^n 2\omega_i \{[y_i(\hat{\theta}_{S_i} - \hat{\theta}_{C_i}) - a(\hat{\theta}_{S_i}) + a(\hat{\theta}_{C_i})]\}$$

A função desvio para o modelo corrente tem a propriedade de aditividade para modelos encaixados.

Quando os modelos não estão encaixados é aplicado o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC).

Critério AIC e BIC

Para um modelo com p parâmetros, o critério AIC é dado por:

$$AIC = Desvio(modelo) + 2p$$

Um outro critério igualmente usual é o BIC que, para um modelo com p parâmetros é dado por:

$$BIC = Desvio(modelo) + 2p \log(n)$$

Ambos os critérios são baseados na função log-verosimilhança com um factor de penalização para o número de parâmetros. Quanto menor for o valor obtido em ambos os critérios melhor será o modelo.

3.3 Modelo de Regressão Logística

3.3.1 Resposta binária

Nesta secção são considerados os modelos lineares generalizados onde a variável resposta é medida numa escala binária. Suponhamos agora que cada indivíduo ou unidade experimental possui apenas duas alternativas na variável resposta, Y , tomando os valores, por convenção, 0 e 1. Uma observação desta natureza, implica uma análise de presença ou ausência de determinada doença, vivo ou morto, a escolha entre duas alternativas com o fim de se testar a importância de cada uma delas, entre outras. De um modo geral, a designação estatística genérica das duas categorias é dada por “sucesso” ($Z = 1$) e “insucesso” ($Z = 0$).

Seja a variável aleatória binária definida por:

$$Z = \begin{cases} 1 & \text{se for sucesso} \\ 0 & \text{se for insucesso} \end{cases}$$

com $P(Z = 1) = \pi$ e $P(Z = 0) = 1 - \pi$. Num inquérito é muito comum existir um número elevado de n observações independentes entre si, para o efeito são consideradas n variáveis aleatórias Z_1, \dots, Z_n independentes entre si, com probabilidades $P(Z_j = 1) = \pi_j$ e a sua distribuição conjunta é dada por:

$$\prod_{j=1}^n \pi_j^{z_j} (1 - \pi_j)^{1-z_j} = \exp \left(\sum_{j=1}^n z_j \log \left(\frac{\pi_j}{1 - \pi_j} \right) + \sum_{j=1}^n \log(1 - \pi_j) \right), \quad z_j = 0, 1$$

onde, por sua vez, faz parte da família exponencial.

De seguida, para o caso onde as probabilidades de sucesso são todas equivalentes vem que:

$$\tilde{Y} = \sum_{j=1}^n Z_j$$

onde \tilde{Y} corresponde ao número de sucessos em n observações. A variável aleatória Y segue, assim, uma distribuição $Bin(n, \pi)$ e a sua função massa de probabilidade é dada por:

$$f(y) = P(\tilde{Y} = \tilde{y}) = \binom{n}{\tilde{y}} \pi^{\tilde{y}} (1 - \pi)^{n - \tilde{y}}, \quad \tilde{y} = 0, 1, \dots, n$$

Por fim, são consideradas N variáveis aleatórias independentes Y_1, Y_2, \dots, Y_N correspondentes ao número de sucessos em N grupos diferentes. Em muitas situações experimentais as respostas podem surgir agrupadas resultante do facto de poder haver mais do que um indivíduo com o mesmo padrão de covariáveis. Desta forma, considera-se a variável $Y_i \sim \text{Bin}(n_i, \pi_i)$ e a sua função log-verosimilhança é dada por:

$$l(\pi_1, \dots, \pi_N; \tilde{y}_1, \dots, \tilde{y}_N) = \left[\sum_{i=1}^N \tilde{y}_i \log \left(\frac{\pi_i}{1 - \pi_i} \right) + n_i \log(1 - \pi_i) + \log \binom{n_i}{\tilde{y}_i} \right]$$

onde n_i é o número de indivíduos para cada combinação de covariáveis, $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$, e \tilde{y}_i corresponde à opção tomada pelo grupo i , isto é, número de sucessos no grupo i , $i = 1, \dots, N$.

	Grupos			
	1	2	...	N
Sucessos	\tilde{Y}_1	\tilde{Y}_2	...	\tilde{Y}_N
Insucessos	$n_1 - \tilde{Y}_1$	$n_2 - \tilde{Y}_2$...	$n_N - \tilde{Y}_N$
Total	n_1	n_2	...	n_N

Tabela 3.1: Proporção de Sucessos/Insucessos versus Grupos de Covariáveis

Tendo em conta que o objetivo principal do estudo é o de estudar a proporção de sucessos em cada grupo caracterizado pela sua combinação de covariáveis. Irá ser considerado como variável resposta uma proporção de sucessos, $Y_i = \frac{\tilde{Y}_i}{n_i}$. Desta forma, vem que $E(\tilde{Y}_i) = n_i \pi_i$ e $E(Y_i) = \pi_i$ sendo esta última a probabilidade pretendida.

A f.m.p. da v.a. Y , omitindo o índice i , é dada por:

$$\begin{aligned} f(y) = P(Y = y) &= \binom{n}{ny} \pi^{ny} (1 - \pi)^{n - ny} = \binom{n}{ny} \left(\frac{\pi}{1 - \pi} \right)^{ny} (1 - \pi)^n \Leftrightarrow \\ &\Leftrightarrow \binom{n}{ny} \left(\frac{\pi}{1 - \pi} \right)^{ny} \left(1 + \frac{\pi}{1 - \pi} \right)^{-n} = \exp \left\{ ny \log \left(\frac{\pi}{1 - \pi} \right) - n \log \left(1 + \frac{\pi}{1 - \pi} \right) + \log \binom{n}{ny} \right\} \Leftrightarrow \\ &\Leftrightarrow \exp \left\{ ny \log \left(\frac{\pi}{1 - \pi} \right) - n \log \left[1 + \exp \left(\log \left(\frac{\pi}{1 - \pi} \right) \right) \right] + \log \binom{n}{ny} \right\} \Leftrightarrow \\ &\Leftrightarrow \exp \{ n(y\theta - a(\theta)) + c(y, \phi) \}, \end{aligned}$$

onde $\theta = \log \left(\frac{\pi}{1 - \pi} \right)$, $a(\theta) = \log(1 + \exp(\theta))$, $\phi = 1$, $c(y, \phi) = \log \binom{n}{ny}$ e $\omega = n$.

Para $y = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1$, pertencendo assim à família exponencial e o valor médio e a variância, com a devida adaptação, são dados por:

$$E(Y) = a'(\theta) = \frac{\exp(\theta)}{1 + \exp(\theta)} = \frac{\frac{\pi}{1 - \pi}}{\frac{1}{1 - \pi}} = \pi$$

e

$$Var(Y) = \frac{\phi}{\omega} a''(\theta) = \frac{1}{n} \frac{\exp(\theta)}{(1 + \exp(\theta))^2} = \frac{1}{n} \frac{\frac{\pi}{1-\pi}}{\frac{1}{(1-\pi)^2}} = \frac{\pi(1-\pi)}{n}$$

3.3.1.1 Função de ligação logit

Com o fim de se investigar a relação entre a probabilidade da variável resposta π_i e o vector de covariáveis $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ é necessário construir um modelo capaz de descrever o efeito de π_i quando o conjunto de covariáveis, \mathbf{x}_i^T , varia. É assumido que essa dependência de π_i sobre \mathbf{x}_i^T ocorre sobre a combinação linear

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j$$

para parâmetros β_1, \dots, β_p desconhecidos. Uma vez que o domínio de η_i é \mathbb{R} , contando que não existam restrições para os parâmetros β_1, \dots, β_p não se pode expressar π_i como combinação linear dos mesmos, pois tal seria inconsistente com as leis de probabilidade. No entanto, existe um método de transformação, $g(\pi_i)$, simples e eficaz de evitar essa dificuldade. Esta solução leva à instância do modelo linear generalizado onde a parte determinística é dada por:

$$g(\pi_i) = \eta_i = \sum_{j=1}^p x_{ij}\beta_j$$

Uma vez que $\theta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$, a função de ligação *logit* ou logística é dada por:

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j=1}^p x_{ij}\beta_j$$

onde π_i é a probabilidade de sucesso, $\pi_i = P(Y = 1 \mid \sum_{j=1}^p x_{ij}\beta_j)$. Esta transformação permite obter uma interpretação mais coerente sobre os parâmetros do modelo de regressão logística.

A probabilidade de sucesso, π_i , é escrita, na sua forma matricial, por:

$$\pi_i = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}$$

3.3.2 Interpretação dos parâmetros

Com vista a obter conclusões de uma análise de uma forma consistente, é fundamental definir uma escala dos efeitos estimados adequada e que seja facilmente interpretada. As escalas consideradas mais adequadas para o modelo de regressão logística são os chamados *odds* e *odds ratio* (OR).

3.3.2.1 Odds

O cálculo do *odds* de um determinado acontecimento baseia-se na razão entre a probabilidade de sucesso desse acontecimento, π_i , e a probabilidade de insucesso do mesmo acontecimento, $1 - \pi_i$.

Supondo que foram usadas apenas duas covariáveis no modelo de regressão logística, nesse caso a sua expressão é dada por:

$$Odds = \frac{\pi_i}{1 - \pi_i} = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$$

Sendo que a probabilidade de sucesso associada, já referido anteriormente, é dada por:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}$$

Ainda de referir que o *odds* pode assumir qualquer valor positivo, ao contrário da probabilidade. Desta forma, se o seu valor for superior a 1 significa que é mais provável obter um sucesso no dado acontecimento do que um insucesso. Por outro lado, se for inferior a 1 a conclusão será a contrária.

3.3.2.2 Odds Ratio

Quanto ao *odds ratio*, é um método baseado na razão entre dois *odds*. Para o caso onde se tem apenas duas categorias na variável resposta, ($Y = 0$ e $Y = 1$), o OR é dado por:

$$OR = \frac{\frac{\pi_1}{(1 - \pi_1)}}{\frac{\pi_2}{(1 - \pi_2)}}$$

Este método tem como principal objetivo comparar ambos os grupos de dados binários, avaliando assim qual o acontecimento com maior propensão de se concretizar. O seu domínio é sempre positivo, dado tratar-se de uma razão entre dois *odds* (também positivos).

Existem 3 casos distintos que podem surgir no cálculo deste método:

- Os *odds* de ambos os acontecimentos são iguais, o $OR = 1$. Sendo que este caso apenas acontece quando as probabilidades de sucesso são iguais para ambos os grupos.
- O *odds* do 1º grupo é inferior ao *odds* do 2º grupo, o $OR < 1$. Este caso revela que o 2º grupo possui uma maior propensão de se vir a concretizar sobre o 1º grupo.
- O $OR > 1$, verifica-se precisamente a situação contrária à descrita no passo anterior.

Se se considerar o logaritmo vem que:

$$\log(OR) = \log\left(\frac{\pi_1}{(1 - \pi_1)}\right) - \log\left(\frac{\pi_2}{(1 - \pi_2)}\right)$$

Correspondendo à diferença entre os *logit* dos dois grupos.

O OR também pode descrever o efeito sobre a escolha independentemente das covariáveis consideradas.

Omitindo o índice de agrupamento, i , tem-se para o caso geral:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

onde $\pi = P(Y = 1 | \mathbf{x}^T)$ com $\mathbf{x}^T = (x_1, \dots, x_k)$ e $p = k + 1$.

Se se pretender calcular a alteração produzida no *logit*, parâmetro β , pela alteração de uma unidade da covariável correspondente, o modelo anterior pode ser adaptado da seguinte forma:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \beta_j x_j + M$$

onde

$$M = \beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \cdots + \beta_k x_k$$

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}$$

De acordo com a nova expressão para $\log\left(\frac{\pi}{1-\pi}\right)$, é verificado que o parâmetro β_j corresponde à alteração produzida no logit pela alteração de uma unidade em x_j desde que:

- x_j seja uma covariável com efeito linear;
- a interação de x_j com as restantes covariáveis seja nula;
- as restantes covariáveis permaneçam constantes.

É possível adaptar a expressão de $\log\left(\frac{\pi}{1-\pi}\right)$ em função do odds, ficando:

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_k x_k) = \exp(\beta_j x_j + M) = \exp(\beta_j x_j) \exp(M)$$

No caso do OR, se for considerada uma alteração de u unidades em x_j vem que:

$$\frac{\exp(\beta_j(x_j + u)) \exp(M)}{\exp(\beta_j x_j) \exp(M)} = \exp(\beta_j u)$$

3.3.3 Inferência sobre o modelo de regressão logística

Uma vez que o modelo de regressão logístico é um caso particular dos modelos lineares generalizados, o estimador para o vector $\boldsymbol{\beta}$ é obtido da mesma forma da secção 3.2.3, assim como os testes de hipóteses e intervalos de confiança também igualmente válidos. Contudo, a interpretação dos parâmetros nos modelos de regressão logística é baseada no cálculo do odds ratio, pelo que é necessário adaptar.

3.3.3.1 Intervalos de Confiança aproximados

Em determinadas condições de regularidade vem que a distribuição assintótica de $\hat{\beta}_j$ é:

$$\hat{\beta}_j \sim N\left(\beta_j, I_{jj}^{-1}(\boldsymbol{\beta})\right)$$

onde $\hat{\beta}_j$ é o estimador da máxima verosimilhança. Tomando por $z_{(1-\alpha/2)}$ o quantil $(1 - \alpha/2)$ da distribuição gaussiana padrão, $N(0,1)$, são obtidos os seguintes intervalos de confiança aproximados ao nível de confiança $(1 - \alpha)$.

Intervalo de confiança aproximado para β_j

$$\left[\hat{\beta}_j \pm z_{(1-\alpha/2)} \sqrt{\text{Var}(\hat{\beta}_j)} \right]$$

Intervalo de confiança aproximado para $\exp(\beta_j)$

$$\left[\exp \left(\hat{\beta}_j \pm z_{(1-\alpha/2)} \sqrt{\text{Var}(\hat{\beta}_j)} \right) \right]$$

Intervalo de confiança aproximado para o logit ($\hat{\eta}_i = g(\hat{\pi}_i)$)

$$\left[g(\hat{\pi}_i) \pm z_{(1-\alpha/2)} \sqrt{\text{Var}(g(\hat{\pi}_i))} \right]$$

com

$$\text{Var}(g(\hat{\pi}_i)) = \sum_{j=1}^p x_{ij}^2 \text{Var}(\hat{\beta}_j) + \sum_{j=1}^p \sum_{k=1, k \neq j}^p x_{ij} x_{ik} \text{Cov}(\hat{\beta}_j, \hat{\beta}_k) = \mathbb{X}_i^T I^{-1}(\hat{\beta}) \mathbb{X}_i$$

Intervalo para π_i

Dado que

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}$$

não é uma função linear de β , a dedução do intervalo de confiança baseia-se no:

Método Delta

A ideia deste método é adoptada em casos onde não se tem apenas uma simples soma linear de observações, criando uma aproximação linear usando o desenvolvimento de Taylor e derivando a variância dessa aproximação.

Seja $G_{ij}(\hat{\beta})$ o vector linha das primeiras derivadas de $\hat{\pi}_i$ em ordem a $\hat{\beta}$. A j -ésima componente do vector é dada por:

$$G_{ij}(\hat{\beta}) = \frac{\partial \hat{\pi}_i}{\partial \hat{\beta}_j} = \frac{x_{ij} \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}, \quad j = 0, \dots, k$$

e a aproximação da variância de $\hat{\pi}_i$ é $G_{ij}(\hat{\beta}) I^{-1}(\hat{\beta}) G_{ij}^T(\hat{\beta})$ e o intervalo de confiança aproximado ao nível $(1 - \alpha)$ para $\hat{\pi}_i$ é dado da seguinte forma:

$$\left[\hat{\pi}_i \pm z_{(1-\alpha/2)} \sqrt{G_{ij}(\hat{\beta}) I^{-1}(\hat{\beta}) G_{ij}^T(\hat{\beta})} \right]$$

3.3.4 Medidas de Qualidade

Uma vez construído o modelo é assumido que, numa fase preliminar da modelação, a sua formulação é satisfatória. Desta forma, o modelo contém todas as variáveis (relevantes e respetivas interacções) que devem estar no modelo e as covariáveis foram introduzidas na

forma estrutural correcta. De seguida, é necessário determinar qual a capacidade do modelo na descrição da variável resposta. Com efeito, os métodos estatísticos incumbidos dessa função são designados de métodos de qualidade de ajustamento.

Quando a construção do modelo está concluída, um conjunto de passos deve ser seguido com o fim de se obter o ajustamento do modelo. As componentes dessa abordagem são as seguintes:

- Cálculo e avaliação de medidas de ajustamento gerais;
- Avaliação de cada componente individual no sumário estatístico e na respetiva representação gráfica;
- Análise de outras medidas associadas à diferença ou distância entre os valores observados e ajustados.

Na regressão linear, as medidas de ajustamento e o respectivo diagnóstico do modelo são funções baseadas nos resíduos, onde estes, por sua vez, são definidos pela diferença entre os valores observados e os valores ajustados. Na regressão logística existem diversos métodos para avaliar a qualidade de ajustamento dos seus modelos, isto é, de analisar a diferença entre os valores observados e ajustados. Entre os métodos disponíveis, os mais usuais são a estatística de *Pearson* e a função de desvio (*Deviance*). Outra estatística que pode também ser utilizada é o coeficiente de correlação (R^2).

3.3.4.1 Estatística χ^2 de Pearson

De salientar que os valores ajustados, nos modelos de regressão logística, são calculados para cada uma das classes de covariáveis e dependem da probabilidade estimada para essa mesma classe. A medida residual de *Pearson* para uma determinada classe de covariáveis é dado por:

$$r(y_i, \hat{\pi}_i) = \frac{(y_i - n_i \hat{\pi}_i)}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$$

onde $n_i \hat{\pi}_i = \hat{y}_i$ é o valor ajustado para a i -ésima classe de covariáveis.

A Estatística de *Pearson* generalizada é dada por:

$$X^2 = \sum_{i=1}^J \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} = \sum_{i=1}^J r^2(y_i, \hat{\pi}_i)$$

Onde I é o número de classes de covariáveis (padrões de covariáveis). A estatística é testada comparando X^2 com o quantil de probabilidade $1 - \alpha$ de uma distribuição qui-quadrado com $I - p - 1$. A estatística de *Pearson* pode ser de má qualidade mesmo em amostras de grandes dimensões.

Para o caso particular onde $n_i = 1$ então $\hat{\pi} = \bar{y}$ e vem que:

$$X^2 = \sum_{i=1}^I \frac{(y_i - \bar{y})^2}{\bar{y}(1 - \bar{y})} = n$$

E por este motivo, não é aconselhável, neste caso, usar a estatística.

3.3.4.2 Função Desvio

A medida de qualidade de ajustamento do modelo para dados binomiais, desvio, é dada por:

$$D = \sum_{i=1}^I d(y_i, \hat{\pi}_i)^2 = 2 \sum_{i=1}^I \left[y_i \ln \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \ln \left(\frac{(n_i - y_i)}{n_i (1 - \hat{\pi}_i)} \right) \right]$$

A distribuição para a estatística de desvio, sob a hipótese de que o modelo ajustado está correctamente construído, é um qui-quadrado com $I - (p + 1)$ graus de liberdade. Esta situação deve-se ao facto de que o desvio é uma estatística de teste baseada na razão de verosimilhança de um modelo saturado com I parâmetros com um modelo ajustado de $p + 1$ parâmetros. A distribuição assintótica do desvio é uma qui-quadrado com $I - p$ graus de liberdade, embora esta aproximação não seja razoável mesmo para dimensões de amostras elevadas.

Desta forma, a análise de desvio deve ser utilizada como um critério de suporte de qualidade do modelo, sendo que um valor pequeno desta estatística é um bom indicador de ajustamento.

Para o caso particular onde $n_i = 1$ a função de desvio toma a seguinte forma:

$$-2 \sum_{i=1}^I \left\{ y_i \log \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) + \log(1 - \hat{\pi}_i) \right\} = -2 \left[\hat{\beta}^T \mathbf{x}^T y + \sum_{i=1}^I \log(1 - \hat{\pi}_i) \right]$$

Devido à sua dependência de $\hat{\beta}$, não se torna aconselhável usar este critério de adequabilidade.

3.3.4.3 Coeficiente de Determinação

No modelo linear o coeficiente de determinação, R^2 , define a proporção de variância da variável resposta de acordo com as variáveis explicativas, sendo que valores elevados desta estatística indicam uma maior proporção de variância explicada por parte do modelo. Pode tomar valores entre -1 e 1 . Nos modelos de regressão logística, com uma variável resposta categórica, não é possível definir um coeficiente de determinação que contenha todas as características provenientes do modelo linear. Para esse efeito, foram definidos alguns métodos alternativos na estimação do coeficiente de determinação para os modelos de regressão logística, entre eles o mais usado é o R^2 de McFadden. Ao conjunto das medidas associadas a esta estatística, para os modelos de regressão logística, dá-se o nome de pseudo-coeficiente de determinação, pseudo- R^2 .

A estatística de *McFadden* é baseada na comparação entre um modelo sem qualquer covariável (modelo nulo) e um modelo com todas as covariáveis disponíveis (modelo saturado). É definido da seguinte forma:

$$R_{McFadden}^2 = 1 - \frac{\log(L(M_N))}{\log(L(M_S))}$$

onde a razão das funções de verosimilhança representa o nível de melhoramento por parte do modelo saturado sobre o modelo nulo. A função de verosimilhança recai num valor entre 0 e 1 , desta forma o seu logaritmo tomará um valor igual ou inferior a 0 . Assim, uma razão de log-verosimilhanças reduzida indica que o modelo saturado produz um melhor ajustamento do que o modelo nulo. O modelo que apresentar o maior valor desta estatística é considerado o melhor. Contudo, os seus valores tendem a ser baixos, mesmo nos modelos onde os seus parâmetros revelam uma forte associação com a variável resposta [Hu, Bo., Shao, Jun. e Palta, Mari. (2006)].

3.3.5 Análise de Resíduos

Após efectuado o ajustamento do modelo de regressão logística aos valores observados, é necessário verificar os pressupostos do modelo relativamente à distribuição da função de

ligação, dos valores preditos assim como os valores mal ajustados. Todos os pressupostos são validados através de uma análise sobre os resíduos.

Existem diversas formas de se calcular os resíduos, todavia a sua finalidade consiste em avaliar a proximidade entre os valores observados e os valores ajustados, assim é possível obter um nível de concordância entre os valores mencionados de acordo com as covariáveis do modelo. A fórmula dos resíduos para a i -ésima observação do conjunto de dados é dada por:

$$r_i = y_i - \hat{y}_i$$

sendo que y_i é o valor observado para a i -ésima classe de covariáveis e \hat{y}_i é o valor ajustado para a i -ésima classe de covariáveis. Onde $\hat{y}_i = n_i \hat{\pi}_i$.

Tal como já foi dito anteriormente, existem diferentes formas de se calcularem os resíduos, sendo apresentados apenas aqueles que foram utilizados no capítulo 5. Para os modelos de regressão existem dois tipos de resíduos mais usuais, nomeadamente os resíduos de *Pearson* e os resíduos de desvio.

3.3.5.1 Resíduos de Pearson

Os resíduos de *Pearson*, ou qui-quadrado, são dados por:

$$X_i = \frac{(y_i - n_i \hat{\pi}_i)}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$$

Nestes casos a variância pode não ser igual a 1, ao contrário do caso dos modelos de regressão linear. Com efeito, usando a aproximação de regressão linear de *Pregibon* (1981) no resíduo para a i -ésima classe de covariáveis, $Var[y_i - n_i \hat{\pi}_i] \approx (1 - h_i) Var(y_i)$, onde h_i é a medida *leverage*. A variância do resíduo é dada por

$$n_i \hat{\pi}_i [1 - \hat{\pi}_i] (1 - h_i)$$

o que sugere que os resíduos de *Pearson* não têm variância igual a 1 excepto se forem padronizados.

Os resíduos de *Pearson* padronizados são dados por:

$$r_{pi} = \frac{(y_i - n_i \hat{\pi}_i)}{\sqrt{n_i \hat{\pi}_i [1 - \hat{\pi}_i] (1 - h_i)}} = \frac{X_i}{\sqrt{(1 - h_i)}}$$

3.3.5.2 Resíduos de Desvio

Estes são desenvolvidos a partir da função desvio, já mencionada anteriormente. A função de desvio residual é definida do seguinte modo:

$$d(y_i, \hat{\pi}_i) = \pm \left\{ 2 \left[y_i \ln \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \ln \left(\frac{(n_i - y_i)}{n_i (1 - \hat{\pi}_i)} \right) \right] \right\}^{1/2}$$

onde o sinal + ou - é o mesmo que o sinal de $(y_i - n_i \hat{\pi}_i) = (y_i - \hat{y}_i)$, isto é, dos resíduos do modelo.

Para as classes de covariáveis com $y_i = 0$ o desvio residual é dado por:

$$d(y_i, \hat{\pi}_i) = -\sqrt{2n_i |\ln(1 - \hat{\pi}_i)|}$$

enquanto o desvio residual para as classes de covariáveis com $y_i = n_i$, é dado por:

$$d(y_i, \hat{\pi}_i) = \sqrt{2n_i |\ln(\hat{\pi}_i)|}$$

Os resíduos de desvio padronizados são dados pela seguinte fórmula:

$$r_{di} = \frac{d(y_i, \hat{\pi}_i)}{\sqrt{(1 - h_i)}}$$

onde é possível verificar que os resíduos de desvio padronizados, tal como os resíduos de *Pearson* padronizados, dependem do *leverage*.

3.3.5.3 Pontos influentes e Leverage

O *Leverage*, h_i , é um termo usado nos modelos de regressão e, em particular, em análises com vista a identificar as observações que se encontram afastadas das restantes de acordo com as covariáveis. As observações *leverage* não têm necessariamente um grande efeito na variável resposta nos modelos de regressão ajustados. Apesar de uma observação influente possuir um *leverage* elevado, um *leverage* elevado não é necessariamente uma observação influente.

Por outro lado, uma observação é influente se a sua exclusão produzir alterações significativas nas estimativas dos parâmetros do modelo e consequentemente no seu ajustamento. A sua presença ou ausência pode, por esse motivo, alterar as conclusões finais do modelo.

3.3.5.4 Outliers

Um *outlier* consiste numa observação que se encontra distante das restantes observações. A existência de *outliers* pode revelar alguma variabilidade no conjunto de dados ou mesmo ser indicador de um erro experimental. De um modo geral, são excluídos do conjunto de dados.

Uma causa frequente dos mesmos é a mistura de duas distribuições, indicando a possibilidade da presença de duas sub-populações (modelo misto).

3.3.5.5 Análise gráfica dos resíduos

Os procedimentos gráficos, na análise de resíduos, são de extrema importância. Após o ajustamento de um modelo ao conjunto de dados é preciso diagnosticar possíveis problemas na parte sistemática e na parte aleatória através de representações gráficas.

Identificação de Outliers

Com vista a identificar quais as observações consideradas *outliers*, existem diferentes tipos de representações gráficas para o efeito.

Gráfico dos resíduos padronizados de Pearson vs Índice das observações

Considerado o gráfico mais eficaz na detecção de *outliers*, dado que permite encontrar os resíduos com valores demasiado elevados. Este gráfico também pode avaliar o ajustamento do modelo através da disposição dos resíduos em função do seu índice de observação, sendo um modelo correcto se este não apresentar uma tendência e revelar uma estabilidade em torno do zero.

Gráfico das probabilidades cruzadas vs Índice das observações

Esta representação é utilizada exclusivamente para dados não agrupados e visa calcular a probabilidade da i -ésima observação sabendo as restantes $n-1$ observações. Esta probabilidade é denominada de probabilidade cruzada e o seu cálculo contra a sua ordem de observação visa saber quais as observações que têm menos probabilidade de acontecer, sendo estas possíveis *outliers*.

Gráfico half-normal dos resíduos padronizados de Pearson

Quando o modelo ajustado é o correcto o papel de probabilidades normal pode ser útil na análise do modelo, sendo que os resíduos de desvio seguem uma distribuição normal padrão sob a condição anterior. Contudo, existem casos onde a sua distribuição não é gaussiana mesmo sendo um modelo correctamente ajustado. Para estes casos, as observações não estão certamente sobre uma linha recta e podem revelar alguma correlação entre si no entanto, este gráfico, assim como o gráfico dos resíduos padronizados de *Pearson* vs índice das observações, pode detectar a presença de *outliers*.

Gráfico dos Leverages vs Índice das observações

Este gráfico consiste em detectar as observações mais influentes para o modelo, isto é, observações que ao serem retiradas do modelo produzam elevadas alterações nos coeficientes de cada uma das covariáveis presentes no modelo. A medida que permite detectar essas observações é denominada de *leverage*, h_i , sendo uma observação é considerada influente se $h_i > 2p/n$, onde p é o número de parâmetros do modelo e n o número de observações.

Gráfico da estatística D vs Índice das observações

Este gráfico permite avaliar a influência da i -ésima observação no vector das estimativas dos parâmetros, obtido com base em todas as observações.

A expressão para a estatística D de cada observação i é dada por:

$$D_i = \frac{h_i r_{pi}^2}{p(1 - h_i)}$$

Valores elevados de D_i revelam uma elevada influência por parte da i -ésima observação, a sua representação contra a sua ordem de observação é um método útil de se detectarem observações influentes. Esta estatística é considerada a mais útil na avaliação da influência produzida por uma observação na estimativa dos parâmetros do modelo (Collett, 1999).

3.3.6 Capacidade preditiva do modelo

A capacidade preditiva do modelo tem como principal objetivo medir a aptidão para discriminar os indivíduos que genuinamente optaram por aquela alternativa de interesse daqueles que não optaram por essa via e preferiram a outra. Apesar disso, não é imperativo que um modelo que se ajusta bem aos dados tenha uma boa capacidade preditiva.

3.3.6.1 Erro de predição

Uma das abordagens mais usuais na análise do erro de predição é a representação de uma tabela de contingência entre as frequências dos valores observados e as frequências dos valores ajustados. Para o efeito, é necessário definir uma nova variável dicotómica que para um ponto de corte igual a 0.5 tomará o valor 1, se $\hat{\pi} > 0.5$, e o valor 0 $\hat{\pi} < 0.5$. Os valores ajustados ficam assim agrupados em duas categorias, permitindo desta forma obter uma ideia da capacidade que o modelo tem em discriminar correctamente entre ambas.

3.3.6.2 Curva ROC

O gráfico relativo à curva ROC produz o mesmo resultado que a tabela de contingência, permitindo identificar os verdadeiros positivos e negativos. O gráfico é construído de acordo com o valor da sensibilidade (eixo vertical) e da falta de especificidade (eixo horizontal) e a medida que avalia a capacidade de discriminação do modelo é dada pela área sob a curva ROC (*AUC-Area under the curve*), sendo que esta pode variar entre 0 e 1. Em qualquer caso estudo, o valor mínimo que a área pode tomar é de 0.5, embora seja apenas considerada uma discriminação aceitável a partir de 0.7. Este gráfico representa a curva ROC e a sua respetiva

área sob a mesma de acordo com o *cutpoint* ótimo, isto é, o seu processamento passa pela determinação do *cutpoint* que produzir uma melhor discriminação possível da variável resposta do modelo em causa. Para além disso, também se encontram representadas as taxas de detecção de falsos positivos e as taxas de omissão de falsos negativos (*false discovery rate* e *false omission rate*), PV+ e PV- respetivamente. Estas duas últimas medidas indicam as percentagens do erro de tipo I e do erro de tipo II, isto é, a capacidade de indiscriminação de acordo com os valores ajustados.

Área sob a curva ROC	Critério
0.5	Modelo sem capacidade discriminativa
$0.7 \leq AUC \leq 0.8$	Modelo com capacidade discriminativa aceitável
$0.8 \leq AUC \leq 0.9$	Modelo com capacidade discriminativa excelente
$AUC \geq 0.9$	Modelo com capacidade discriminativa extraordinária

Tabela 3.2: Critérios de discriminação

Capítulo 4

Análise Exploratória dos dados

No presente capítulo será feita uma breve descrição sobre a população e a amostra em estudo, referentes ao município de Águeda, seguido de uma análise exploratória sobre os dados do inquérito.

4.1 A população e a amostra

Com o objetivo de se proceder a uma análise exploratória dos dados foi obtida uma amostra representativa da população em questão. Este procedimento torna-se necessário devido à elevada dimensão da população em causa, de outra forma a análise seria efetuada sobre um conjunto de dados global, implicando um gasto desnecessário em recursos. Por outro lado, se a dimensão de amostra for demasiado reduzida, a representatividade dos dados diminui.

Para o efeito, foi estimado o erro máximo incorrido, dada a dimensão da amostra obtida no inquérito e a dimensão da população de onde foi retirada essa amostra, visando deste modo obter uma precisão segura quanto à sua representatividade. O cálculo desse erro foi baseado nas propriedades da estimação da variância de uma proporção.

A variância de uma proporção é dada pela seguinte expressão:

$$V(p) = E(p - P)^2 = \frac{S^2}{n-1} \left(\frac{N-n}{N} \right) = \frac{P(1-P)}{n} \left(\frac{N-n}{N-1} \right) \quad (4.1)$$

Onde P é a proporção da população com uma determinada característica, N é a dimensão da população e $S^2 = \frac{N}{N-1} P(1-P)$. Uma vez que P é desconhecido, S^2 também o será. Um estimador centrado de S^2 é $s^2(p) = \frac{N-n}{N} \frac{p(1-p)}{n-1}$. Se $\frac{N-n}{N}$ for negligenciável, o estimador de $V(p)$ toma a forma simplificada

$$s^2(p) = \frac{p(1-p)}{n-1}$$

Observando a expressão (4.1), é possível verificar que esta varia conforme o valor da proporção, tomando o seu valor máximo para $P = (1 - P) = 50\%$, isto é, quando a população é repartida em dois grupos de igual dimensão. Quando a dimensão da população, N , é muito elevada e afastada relativamente à dimensão da amostra, n , $\left(\frac{N-n}{N-1}\right) \approx 1$, podendo ser, neste caso, excluído da expressão.

Nesse caso, a variância é dada pela seguinte expressão:

$$V(p) = \frac{P(1-P)}{n} \quad (4.2)$$

sendo o erro máximo absoluto calculado, aplicando a raiz quadrada sobre a equação 4.2 e d é a tolerância definida para o erro absoluto. De seguida é ilustrado um exemplo com os valores da variância, $P(1 - P)$, e do desvio padrão, $\sqrt{P(1 - P)}$, de uma amostra de dimensão 1, visando observar a pequena variação que ocorre quando a proporção da população varia entre os 30% e os 70% e a necessidade de uma maior dimensão da amostra para o caso onde a proporção da população é igual a 50% (pior situação já referida).

P - proporção (em percentagem) na população

P	0	10	20	30	40	50	60	70	80	90	100
$P(1 - P)$	0	900	1600	2100	2400	2500	2400	2100	1600	900	0
$\sqrt{P(1 - P)}$	0	30	40	46	49	50	49	46	40	30	0

Tabela 4.1: Valores de $P(1 - P)$ e de $\sqrt{P(1 - P)}$

Onde o erro padrão absoluto incorrido, sendo d a tolerância, é dado pela seguinte expressão:

$$\sigma_{abs} = \sqrt{\frac{P(1-P)}{n}} \leq d \quad (4.3)$$

Para o caso onde $\sqrt{(1 - P)P} = 50$, é precisa uma amostra de dimensão 100 a fim de obter uma tolerância para o erro absoluto inferior ou igual a 5%. Para obter uma tolerância igual ou inferior a 1%, é necessária uma amostra de dimensão 2500.

No caso presente a dimensão da amostra é $n=2232$, assim tomando $P = \frac{1}{2}$, o erro absoluto é inferior ou igual a 0.0106.

Intervalo de Confiança para a proporção da população

A distribuição exacta do número X de elementos de uma amostra com uma determinada característica, é *hipergeométrica*, tendo-se

$$P(X = x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}, \max(0, n - N + K) \leq x \leq \min(K, n) \quad (4.4)$$

No caso da dimensão da amostra ser razoavelmente grande, podemos aproximar a distribuição hipergeométrica à binomial, pois o facto da amostra ser extraída sem reposição deixa de ter

relevância. No entanto esta aproximação também não traz grandes vantagens ao cálculo, sendo mais conveniente usar a aproximação da binomial à Gaussiana.

Assim, tendo em conta que a expressão (4.1) é equivalente a

$$V(p) = \frac{P(1-P)}{n} \left(\frac{N-n}{N-1} \right) \frac{N}{N} = \left(\frac{N-n}{N} \right) \frac{N}{N-1} \frac{P(1-P)}{n}$$

e omitindo o factor $\frac{N}{N-1}$, o que é admissível neste contexto, por termos de assumir que a dimensão N da população tem de grande para se poder usar a aproximação à Gaussiana, vem:

$$p \sim \text{Gau} \left(P, \sqrt{\frac{N-n}{N} \frac{P(1-P)}{n}} \right) \quad (4.5)$$

Esta aproximação à Gaussiana permite construir Intervalos de Confiança assintóticos, pois

$$P \left(\frac{|P-p|}{\sqrt{\frac{N-n}{N} \frac{P(1-P)}{n}}} < z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha \quad (4.6)$$

Dado que a variância de P é desconhecida, teremos de usar em (4.5) o seu estimador centrado $s(p)^2 = \frac{N-n}{N} \frac{p(1-p)}{n-1}$, obtendo-se o seguinte intervalo assintótico de confiança para P

$$p \pm z_{1-\alpha/2} \sqrt{\frac{N-n}{N} \frac{p(1-p)}{n-1}} \quad (4.7)$$

onde N é a dimensão da população, n é a dimensão da amostra e $z_{1-\alpha/2}$ o quantil empírico da distribuição normal [Barnett, V. 1974 (45)]. A semi-amplitude d do intervalo de confiança, é um majorante do erro de estimação.

$$d = z_{1-\alpha/2} \sqrt{\frac{N-n}{N} \frac{p(1-p)}{n-1}}$$

Assim, calcular o erro padrão, para uma dimensão amostral fixa, é o mesmo que assegurar que a probabilidade do erro de estimação ultrapassar d é inferior ou igual a α , isto é,

$$P\{|\hat{P} - P| > d\} \leq \alpha$$

De acordo com os dados obtidos pelo inquérito, foram retiradas 2232 observações para o caso estudo. Uma vez que o inquérito se realizou no município de Águeda, foram obtidos os dados referentes à dimensão populacional, desde 1981, nesse mesmo município através do INE (Instituto Nacional de Estatística), nos Censos 2011. Dado que o inquérito foi efectuado no ano de 2013, foi necessário desenvolver uma série temporal com as dimensões populacionais anuais, desde 1981 até 2011, com vista a prever as dimensões populacionais para o ano de 2012 e de 2013.

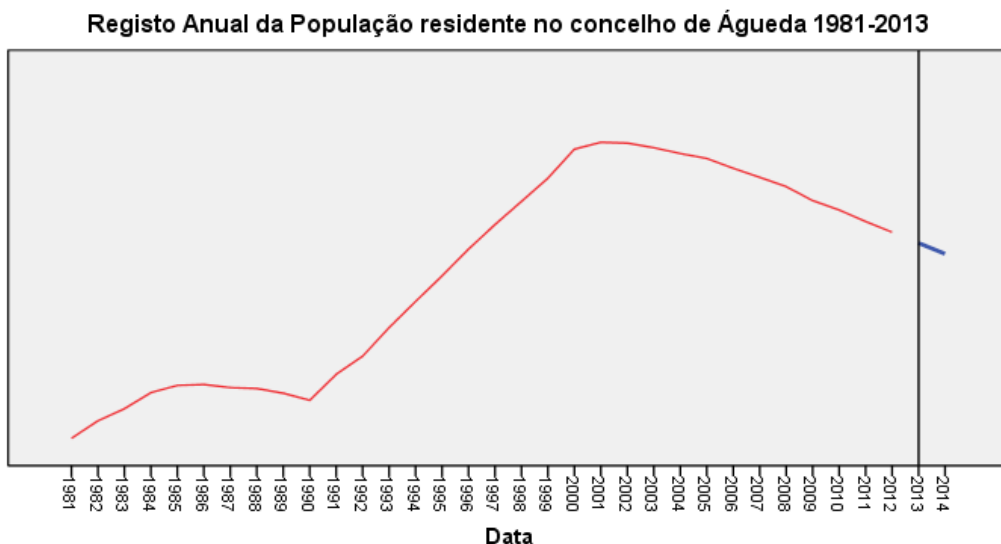


Figura 4.1: Série Temporal relativo às populações anuais no município de Águeda [1981-2014]

De acordo com o gráfico 4.1 verificou-se um aumento significativo da população até ao ano de 2001, havendo uma tendência aditiva até esta data, e a partir de 2011 até à atualidade a população tem vindo a sofrer uma queda gradual na ordem dos 2% a 3%. A previsão para o ano de 2013 e 2014 é que essa tendência decrescente continue. Dado o decréscimo visível nos últimos 10 anos, é possível deduzir o conjunto de valores para o intervalo desde 2001 até 2013, deste modo já será possível observar uma tendência subjacente aos dados populacionais do concelho.

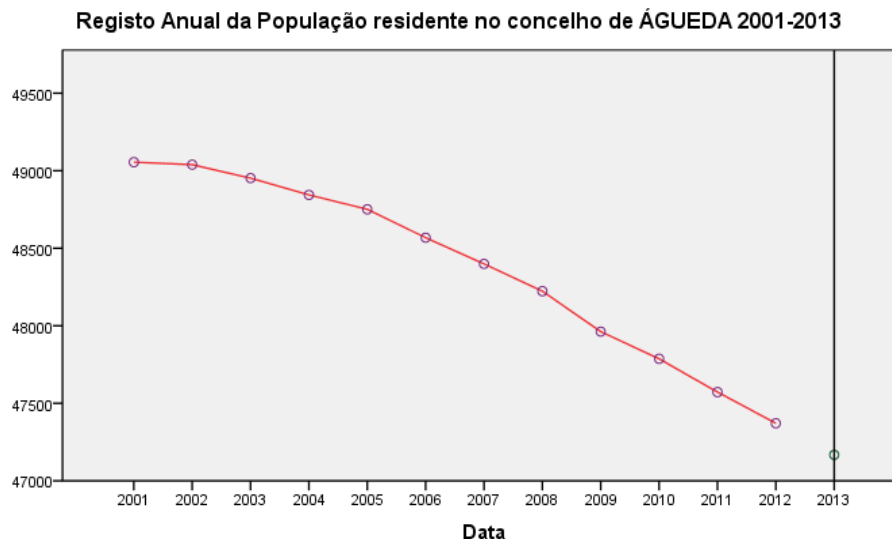


Figura 4.2: Série Temporal relativo às populações anuais no município de Águeda [2001-2013]

Como já havia sido previsto, nos últimos 10 anos já é possível observar uma tendência, na série temporal, no sentido decrescente, sendo que a dimensão populacional prevista para o ano 2013 é de 47167 habitantes.

4.1.1 Erro máximo absoluto

Com base na expressão (4.7) é possível, fixando à partida a dimensão amostral obtida, proceder ao cálculo do erro absoluto cometido na escolha dessa dimensão. Para tal, considera-se um nível de confiança (ou de fiabilidade) de 95%, assumindo desta forma uma margem de erro de 5%, isto é, uma probabilidade de se errar na precisão do cálculo do erro em cerca de 0,05. Da mesma maneira que se pode estimar a dimensão ideal de uma amostra, com uma dada característica, de modo a que esta seja representativa dessa mesma característica com uma dada precisão, também se pode calcular o erro que se toma quando se obtém uma determinada dimensão de entrevistados para a amostra.

Uma vez que não se tem qualquer informação sobre uma média ou um número total de habitantes com uma dada característica da amostra, vai-se tomar uma proporção populacional adequada visando testar a margem de erro incorrida considerando $n=2232$.

Variância

$s^2(p) = (1 - f) * \frac{p(1-p)}{n-1}$, sendo $f = \frac{n}{N}$ (índice de amostragem) e p a proporção amostral. Como $p(1 - p)$ atinge o valor máximo em $\frac{1}{4}$ para $p = \frac{1}{2}$, o pior caso ocorre para este valor de p .

Tomando $N=47169$, $n=2232$ e $p=0,5$, $s^2(p) = 0,000106755$.

Erro máximo incorrido

$$d = z_{1-\alpha/2} * \sqrt{\text{var}(p)} = z_{1-\alpha/2} * \sqrt{s^2(p)} = 1,96 * \sqrt{0,000106755} = 0,020251164$$

Ao se ter entrevistado 2232 pessoas para o estudo incorre-se, no máximo, num erro de cerca de 2%. Uma vez que o cálculo foi feito assumindo um nível de confiança a 95%, ou seja, a um nível de significância de 5%, significa que é possível afirmar que uma amostra de 2232 elementos é suficiente para uma boa representatividade do estudo em causa.

4.2 Tabelas de contingência

4.2.1 Análise de Independência/Associação

Num inquérito torna-se necessário proceder a uma análise sobre as variáveis relativas aos dados pessoais dos inquiridos e sobre as variáveis descritivas das respostas dadas pelos mesmos. Para o efeito, são efetuadas representações gráficas, adequadas ao tipo de valores de cada variável, e posteriormente tabelas de contingência com o intuito de testar uma possível associação entre duas ou mais variáveis qualitativas. A situação onde se tem uma tabela 2 x 2 é o caso mais simples que se pode ter, no entanto quando uma ou ambas as variáveis qualitativas possuem múltiplas categorias, em vez de categorias dicotómicas, a tabela teria mais células do que apenas as 4 no caso 2 x 2. As ditas células representam as frequências cruzadas das categorias, podendo ser transformadas em proporções (obter-se-ia uma interpretação semelhante).

4.2.2 Tabela de Contingência 2 x 2

A forma geral de uma tabela de contingência 2 x 2 é representada na tabela 1, onde uma amostra de N observações é classificada por duas variáveis qualitativas, uma delas contendo r categorias e a outra k categorias. Neste caso, tem-se uma tabela de contingência $r \times k$.

A frequência observada na categoria i da variável linha e na categoria j da variável coluna, isto é, a frequência absoluta da célula ij é representada por n_{ij} . A frequência total, por linha, é denominada por $n_{i.}$ e a frequência total, por coluna, é representada por $n_{.j}$. Estes totais são os chamados totais marginais e são representados em função de n_{ij} da seguinte forma:

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{ik} = \sum_{j=1}^k n_{ij}$$

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{rj} = \sum_{i=1}^r n_{ij}$$

$$n_{..} = \sum_{i=1}^r \sum_{j=1}^k n_{ij}$$

Onde $n_{..}$ representa o número total de observações da amostra relativa ao cruzamento das duas variáveis qualitativas que geralmente é representado por N .

	2ª Variável							
		1	2	.	.	.	k	Total
1ª Variável	1	n_{11}	n_{12}	.	.	.	n_{1k}	$n_{1.}$
	2	n_{21}	n_{22}	.	.	.	n_{2k}	$n_{2.}$

	r	n_{r1}	n_{r2}	.	.	.	n_{rk}	$n_{r.}$
Total		$n_{.1}$	$n_{.2}$.	.	.	$n_{.k}$	$n_{..} = N$

Tabela 4.3: Tabela de Contingência 2 x 2

Como se está a retratar o caso específico de uma tabela de contingência 2 x 2 tem-se $r = k = 2$, ou seja, ambas as variáveis são dicotômicas.

Classificações Independentes - Associação

De um modo geral, o aspeto mais importante a ser analisado é saber se as variáveis qualitativas, que compõe a tabela de contingência, são independentes uma da outra ou não. No caso em que se tem uma tabela 2 x 2 esse tipo de análise torna-se fácil, por exemplo se se quiser cruzar as variáveis género e número de mortos por tuberculose, se estas duas variáveis fossem independentes esperar-se-ia obter uma proporção de mortes por tuberculose no grupo feminino muito semelhante à proporção de mortes no grupo masculino pela mesma causa. Caso exista uma disparidade muito grande destas proporções, a morte por tuberculose tem uma maior tendência de se associar com um dos géneros do que com o outro. O critério de se considerar uma enorme disparidade muitas vezes depende do tipo de valores que essa variável toma, para contornar esse problema é preciso aplicar o chamado teste do qui-quadrado que vai ser discutido de seguida.

Teste Qui-quadrado

Para testar a independência de duas variáveis é preciso investigar a veracidade da hipótese $p_{ij} = p_{i.}p_{.j}$, isto é, que a probabilidade associada à célula da categoria i da variável linha e da categoria j da variável coluna, assumindo a independência, é igual ao produto das probabilidades da categoria i da variável linha e da categoria j da variável coluna. Os estimadores das probabilidades da categoria i da variável linha e da categoria j da variável coluna podem ser calculados, respetivamente, pelas fórmulas $\hat{p}_{i.} = \frac{n_{i.}}{N}$ e $\hat{p}_{.j} = \frac{n_{.j}}{N}$ o que permite estimar a frequência esperada na célula ij da tabela caso as duas variáveis sejam independentes. A hipótese que assume a independência entre as variáveis é denominada de hipótese nula e representa-se por H_0 .

Este teste baseia-se na diferença entre os valores esperados das frequências estimadas (E_{ij}), assumindo a hipótese H_0 verdadeira, e as frequências observadas (n_{ij}). O teste de independência, primeiramente sugerido por *Pearson*, baseia-se na estatística χ^2 definida da seguinte forma:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

onde $E_{ij} = N\hat{p}_{i.}\hat{p}_{.j} = N \frac{n_{i.}n_{.j}}{N} = \frac{n_{i.}n_{.j}}{N}$ (frequência esperada para a célula ij) e n_{ij} a frequência observada na tabela de contingência na linha i e coluna j . Se ambas as variáveis forem independentes, espera-se que a diferença entre a frequência observada e a frequência esperada da célula ij produza um valor muito pequeno caso contrário a diferença terá que ser significativa e, desta forma, as variáveis estarão associadas entre si. Por outras palavras, sob a validade da hipótese H_0 , a estatística χ^2 deverá tomar um valor menor do que quando esta é falsa. Consequentemente, torna-se necessário definir um critério que permita decidir para que valores da χ^2 se rejeita H_0 . Sob a validade da hipótese nula, a estatística χ^2 segue assintoticamente uma distribuição qui-quadrado com $(r-1)(k-1)$ graus de liberdade, logo a hipótese nula será rejeitada ao nível de significância α se

$$\chi^2 > \chi_{1-\alpha; \nu}$$

onde $\nu = (r - 1)(k - 1)$.

Teste Exato de Fisher

Um dos pressupostos para se poder concluir que a distribuição assintótica da estatística χ^2 , sob a validade da hipótese nula, é qui-quadrado, é que as frequências esperadas E_{ij} não sejam muito pequenas. Quando estas frequências são muito pequenas, o teste Exacto de Fisher produz resultados mais eficientes e fiáveis. Este teste, em vez de usar a aproximação ao qui-quadrado, usa a distribuição exacta das frequências observadas. Para totais marginais fixos a distribuição requerida é a chamada distribuição Hipergeométrica, uma vez que está associada a amostragens sem reposição e proveniente de populações finitas. Assumindo a independência das duas variáveis, a probabilidade de se obter qualquer arranjo das frequências a, b, c e d , quando os totais marginais são fixos é:

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!N!}$$

Através da expressão anterior é calculada a probabilidade de se obter a tabela observada, ou a de qualquer outra combinação das frequências que mostre uma evidência igual ou maior da associação entre as variáveis. A soma destas probabilidades é então comparada com o nível de significância α escolhido; se for superior a α conclui-se que a hipótese nula não deve ser rejeitada, caso contrário existem evidências para afirmar que existe associação entre as duas variáveis.

4.2.3 Tabela de Contingência $r \times k$

Ao contrário da fácil interpretação que se pode reter com as tabelas de contingência 2×2 com os testes do qui-quadrado ou com os testes exatos de Fisher; nos casos onde r , o número de linhas, ou k , o número de colunas, ou mesmo ambos são maiores do que 2, as conclusões já não são tão claras apenas com a abordagem que foi referida até ao momento. Quando se têm variáveis com múltiplas categorias, irá ser precisa uma análise mais detalhada a fim de se decidir onde ou se surgirá algum caso de independência. Por outro lado, existe também o problema das categorias ordinais, que não era encontrado nas tabelas 2×2 , apresente-se como exemplo uma análise da gravidade de uma doença definida pelos níveis “baixo”, “médio” e “alto”. Para este tipo de variáveis é igualmente necessário um especial cuidado. Para esta situação, o teste do qui-quadrado é aplicado de forma semelhante ao caso da tabela de contingência 2×2 com exceção na conclusão, onde, em vez de se consultar o valor p , o valor do qui-quadrado, obtido através do teste, é comparado com o quantil da mesma.

Distribuição Qui-quadrado

A distribuição qui-quadrado pode definir-se como sendo a distribuição da soma dos quadrados de um número de variáveis independentes z_i , onde cada uma dessas variáveis segue uma distribuição normal padrão. A forma desta distribuição depende do número de variáveis independentes envolvidas, os chamados graus de liberdade. Isto é,

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_v^2$$

Onde χ^2 tem uma distribuição que apenas depende de v . O número de variáveis independentes envolvidas na distribuição fornece toda a informação necessária para decidir se o valor da estatística χ^2 leva a rejeitar ou não a hipótese de independência. Para tal, é preciso determinar v , onde este depende do número de categorias de cada variável que compõe a tabela de contingência. É possível calcular o número de graus de liberdade através da fórmula:

$$g.l. = rk - r - (k - 1) = (r - 1)(k - 1)$$

Onde r e k correspondem ao número de linhas e ao número de colunas da tabela, respectivamente.

Teste Qui-quadrado

No caso das dimensões da tabela de contingência serem superiores a 2 o critério de se concluir independência ou associação entre as variáveis consiste em comparar o valor observado da estatística χ^2 com o quantil da distribuição qui-quadrado com $= (r - 1)(k - 1)$ g.l., após fixado o nível de significância. Se χ^2 superar o valor tabelado a hipótese de independência tem de ser rejeitada, caso contrário assume-se a independência das variáveis.

Frequências Esperadas reduzidas

Quando as frequências esperadas são muito pequenas (inferiores a 5), o teste do qui-quadrado permanece inválido para as tabelas de contingência de maior dimensão. *Cochran* (1954) argumentou que a regra era demasiado severa e sugeriu que se houver apenas um número reduzido de frequências abaixo de 5 (cerca de 20%), a interpretação mantém-se a cargo do qui-quadrado. No entanto, esta regra pode ser também considerada demasiado restritiva dado a recente investigação de *Lewontin* e *Felsenstein* (1965), *Slakter* (1966), e de outros, mostrando que muitos das frequências esperadas podem tomar valores inferiores a um sem afetar o teste razoavelmente. *Lewontin* e *Felsenstein* apontam ainda que na maioria dos casos, o critério do qui-quadrado pode ser usado nas tabelas com valores esperados em excesso de 0.5 na menor célula.

Um procedimento que tem sido usado por muitos anos com vista a superar o problema das frequências esperadas baixas é o agrupamento das categorias das variáveis. No entanto, pode ser uma ação algo crítica por três razões importantes; primeiro perde-se uma quantidade significativa de informação ao concatenar-se categorias e consequentemente o objetivo do estudo é deteriorado, segundo pode ser afetada a aleatoriedade da amostra (todo o processo envolvendo o teste do qui-quadrado incide sobre a aleatoriedade das amostras) e as categorias em que as observações podem cair são escolhidas antecipadamente. Juntar categorias após ter-se consultado os dados da amostra pode afetar o processo aleatório da mesma, por último a forma como são reunidas as categorias pode ter um efeito importante nas inferências. A prática de se combinar categorias classificativas deve, deste modo, ser evitado se possível.

Isolamento de fontes de associação em tabelas $r \times k$

Um teste qui-quadrado global para uma tabela de contingência $r \times k$ indica a dependência entre duas variáveis, mas não informa de como é que essa associação ocorre ou se existe apenas numa parte da tabela. Para tal, existem métodos que permitem comparar células dentro da tabela. O teste qui-quadrado sobre uma tabela de contingência pode ser particionado em tantas componentes quanto o número de graus de liberdade da tabela. Cada uma dessas componentes corresponde a uma tabela 2×2 construída a partir da tabela original e cada componente é independente uma das outras. Uma vez tendo uma tabela de contingência 2×2 , é possível efectuar o teste do qui-quadrado sobre a mesma e tirar conclusões sobre a partição [*Lancaster and Irwin*, 1949].

4.2.4 Interpretação

Numa fase preliminar do projeto, é necessário um tratamento de dados com o fim de se obterem associações entre variáveis consideradas importantes e influentes no caso estudo, isto é, visando definir grupos distintos e que possam ser representativos na análise. Neste tipo de situações é comum efetuar uma análise com base nas tabelas de contingência, onde estas nos informam sobre as diferenças (existência de associação) ou semelhanças (ausência de associação) entre variáveis que façam sentido de serem comparadas. Nas tabelas são comparadas as frequências tabeladas (dados reais) com as frequências esperadas (dados estimados), são também apresentados os resíduos ajustados que quantificam o grau de associação existente entre as variáveis envolvidas. Em paralelo, são efetuados dois testes, igualmente criteriosos quanto à existência ou ausência de associação, entre eles; o teste qui-quadrado e o teste exacto de Fisher.

A interpretação de cada variável, envolvida e considerada importante no estudo, foi feita através da representação de tabelas de contingência, visando obter possíveis associações (grupos) entre as variáveis de interesse e as respetivas questões disponibilizadas no inquérito. Os critérios de uma possível associação entre as variáveis basearam-se numa análise dos resíduos ajustados, representados na tabela já mencionada, seguido de um teste de hipóteses, avaliando a independência das mesmas.

4.3 Análise Estatística

4.3.1 Simplificar a mobilidade em duas rodas

4.3.1.1 *Qual o modo de transporte que utiliza diariamente na sua deslocação casa/escola?*

Houve 248 estudantes do ensino secundário que responderam ao inquérito. A primeira pergunta é relativa ao tipo de transporte utilizado, diariamente, na sua deslocação casa/escola. Entre as respostas alternativas à pergunta constavam o autocarro escolar, o automóvel, a bicicleta convencional, o comboio e a opção de outro meio de transporte ainda não mencionado. Havendo apenas 9 pessoas que não responderam à questão e 10 a utilizar mais do que um transporte, cerca de 26% usa o autocarro escolar, 51% utiliza o automóvel, apenas 13% opta por ir a pé, nenhum estudante usa a bicicleta convencional, cerca de 10% escolhe viajar de comboio e apenas um estudante opta por um outro transporte.

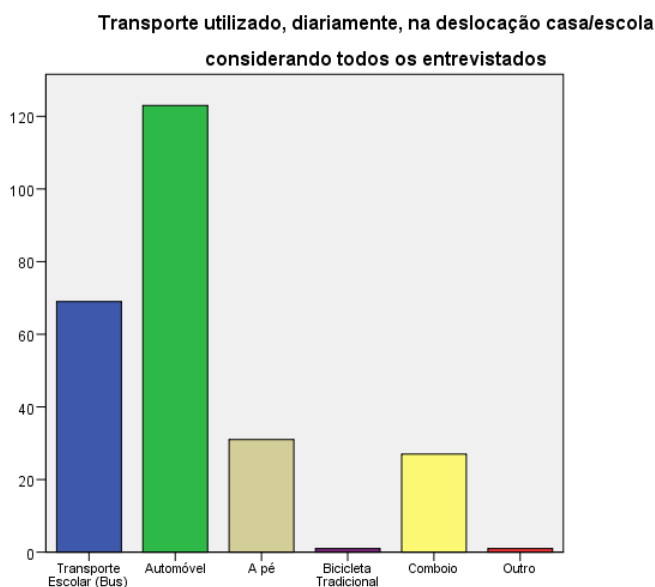


Figura 4.3: Meio de transporte utilizado pelos estudantes

4.3.1.2 Sabe andar de bicicleta?

Do total de inquiridos, 99% afirmaram saber andar de bicicleta.

4.3.1.3 Costuma andar regularmente (todos os dias) de bicicleta convencional?

Apenas os estudantes que afirmaram saber andar de bicicleta puderam responder a esta questão. Relativamente à regularidade no uso da bicicleta convencional, apenas 16% responderam afirmativamente à pergunta.

4.3.1.4 Já utilizou a bicicleta elétrica de Águeda?

Quanto à utilização da bicicleta elétrica de Águeda, houve apenas 6% (apenas 15 estudantes) dos inquiridos a responder afirmativamente à pergunta.

4.3.1.5 Perceção de barreiras na sua deslocação casa/escola, em bicicleta elétrica: De acordo com a sua perceção, que barreiras (obstáculos) existem?

Para a presente pergunta foram consideradas 16 barreiras consideradas influentes no uso da bicicleta elétrica na deslocação casa/escola, sendo pedido aos inquiridos que as ordenassem de acordo com o seu nível de importância. Sendo o número 1 atribuído à barreira de maior importância, 2 à de segunda maior importância e assim sucessivamente [questão exclusiva para os estudantes que responderam afirmativamente à questão anterior].

	1ª	2ª	3ª	Total
Condições Climáticas	8	1	1	10
Pavimento Irregular	6	4	0	10
Tráfego motorizado na infraestrutura (VLP)	5	2	2	9
Ruído de Tráfego	5	2	2	9
Risco de Acidente (com outros veículos)	5	3	0	8
Velocidade de Tráfego	5	3	0	8
Risco de Acidente (com outros peões)	5	2	0	7

Tabela 4.4: Níveis de Importância das barreiras relativamente à bicicleta elétrica

A barreira considerada de 1ª importância, pela maioria dos inquiridos, foi a barreira relativa às condições climáticas, com um total de 8 inquiridos (cerca de 53%). Em 2º lugar encontrou-se a barreira associada ao pavimento irregular, com um total de 6 inquiridos (cerca de 40%). Com uma frequência absoluta de 5 inquiridos, foram obtidos os obstáculos “Tráfego motorizado na infraestrutura (VLP)”, “Ruído de Tráfego”, “Risco de Acidente (com outros veículos)”, “Velocidade de Tráfego nas estradas/ruas” e “Risco de Acidente (com peões)”, ocupando a 3ª posição.

4.3.1.6 Percepção de barreiras na sua deslocação casa/escola, em bicicleta convencional: De acordo com a sua percepção, que barreiras (obstáculos) existem?

Foi feita novamente a mesma pergunta relativa aos 16 obstáculos mas, desta vez, de acordo com o uso da bicicleta convencional. Para esta questão já se obteve um total de 200 respostas, sendo que a barreira de maior importância foi o risco de acidente com outros veículos (com um total, na soma das 3 primeiras categorias, de 116 afirmações), seguido do relevo (com um total de 105 respostas) e, na 3ª posição de importância, as condições climatéricas (com um total de 102 respostas). Para cada uma das barreiras mencionadas, a sua categoria de 1ª maior importância teve um total de 67, 61 e 50 para o risco de acidente com outros veículos, o relevo e as condições climatéricas, respetivamente.

	1ª	2ª	3ª	Total
Risco de Acidente com outros veículos	67	36	13	116
Relevo	61	34	10	105
Condições Climatéricas	50	35	17	102

Tabela 4.5: Níveis de importância das barreiras relativamente à bicicleta convencional

4.3.1.7 Características da bicicleta elétrica: Quais dos seguintes atributos considera mais importantes na bicicleta elétrica?

Foi efetuada uma avaliação às principais características consideradas importantes na bicicleta elétrica, onde foram obtidas 218 respostas válidas. As características que foram mais valorizadas, pela maioria dos inquiridos, foram; o preço acessível (com uma frequência total, no conjunto das 3 primeiras importâncias, de 172 respostas), o esforço físico reduzido (com um total de 106 respostas) e a segurança rodoviária (com um total de 101 respostas). Sendo que, na categoria de 1ª importância, o atributo relativo ao “preço acessível” arrecadou 131 respostas e os restantes dois, já mencionados anteriormente, com um total de 57 cada um.

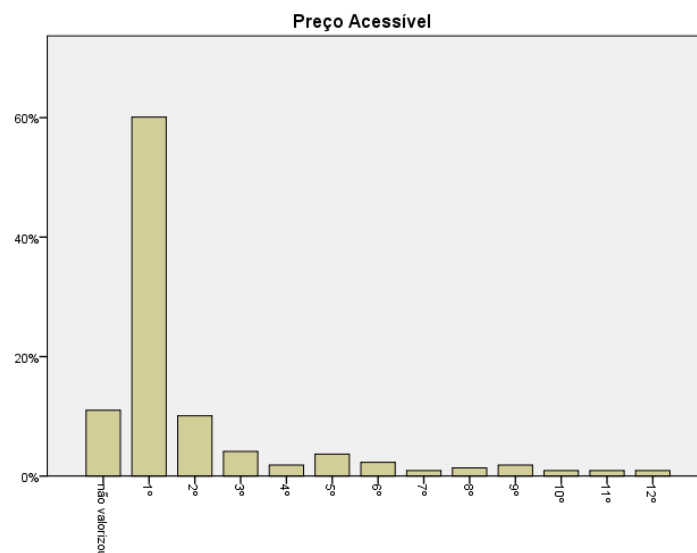


Figura 4.4: Atributo da bicicleta elétrica mais importante por parte dos estudantes

4.3.2 Avaliação das preferências dos inquiridos de acordo com o regime de utilização

4.3.2.1 Regime de utilização da bicicleta elétrica

Do ponto de vista global, relativamente ao regime de utilização da bicicleta elétrica, os inquiridos preferem optar pela utilização da bicicleta convencional em alternativa à compra ou aluguer da bicicleta elétrica. Foram obtidas 232 respostas. Cerca de 50% (115 pessoas) dos inquiridos escolheram a bicicleta tradicional, aproximadamente 17% (40 pessoas) optou pelo aluguer da bicicleta elétrica e cerca de 14% (32 pessoas) optou pela sua aquisição. De destacar que houve ainda cerca de 19% (45 pessoas) que não optou por nenhuma das respostas alternativas, apresentadas anteriormente.

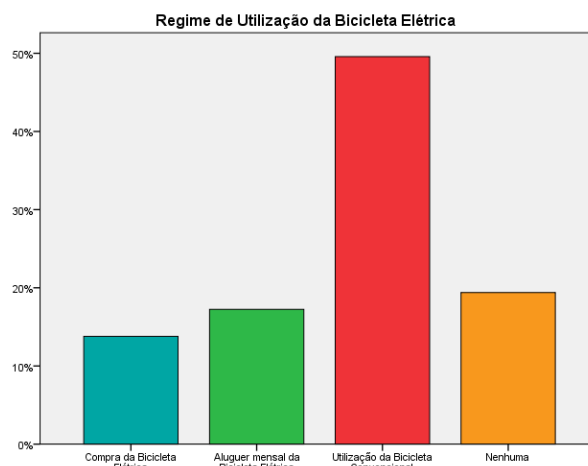


Figura 4.5: Regime de utilização da bicicleta elétrica de maior importância por parte dos estudantes

4.3.2.2 Modelo da bicicleta elétrica

Relativamente à preferência do modelo de bicicleta elétrica a usar, houve um total de 208 respostas válidas. O modelo “Montanha/Desporto” foi o predileto da maioria, com um total de 107 inquiridos (aproximadamente 52%), seguido do modelo “Utilitário”, com um total de 57 pessoas (cerca de 26%), na 3ª posição ficou o modelo “Design personalizado”, com um total de 43 entrevistados (sensivelmente 21%), e, por último, o modelo dobrável com apenas 3 intervenientes (cerca de 1,4%). Houve apenas 1 inquirido a optar por outro modelo, sem ser aqueles já mencionados.

4.3.2.3 Inserção de um sistema de equipamento de novas tecnologias TIC

Dos 239 intervenientes, 123 (51%) afirmaram preferir adicionar as novas tecnologias TIC nos modelos mencionados anteriormente. Em relação à preferência de um outro veículo, de 239 inquiridos, surgiram apenas 34 que manifestaram preferência no uso de outro veículo. Dos 34, 18 usam o carro, 3 usam o carro/comboio e 2 usam a mota.

4.3.2.4 Equipamento adicional

Apenas os inquiridos que responderam afirmativamente à questão anterior puderam responder sobre as suas preferências no equipamento adicional (123). Entre os equipamentos adicionais disponíveis para a bicicleta elétrica, aquele que obteve um maior número de votos foi o sistema de tecnologia de informação e comunicação (com GPS integrado com vista a permitir o cálculo do caminho mais curto a ser percorrido) com um total, na soma dos 3 primeiros níveis de importância, de 79 respostas, de seguida foi o sistema de contabilização de calorias, com um total de 65 votos, e logo depois foi a opção de considerar todos os 4 equipamentos adicionais disponibilizados no questionário (Sistema de Tecnologia de Informação e Comunicação, Sistema de Contabilização de Calorias consumidas na deslocação, Sistema de Tecnologia Bluetooth e o Sistema Cooperativo de Bicicletas), com um total de 61

afirmações. Sendo que foram obtidas várias respostas onde foi dada a mesma importância a mais do que um equipamento, tendo em atenção apenas a categoria de 1ª importância, foram obtidos 62, 53 e 31 votos (50% para o sistema de TIC, 43% para os 4 equipamentos em simultâneo e 25% para o sistema de contabilização de calorias, respetivamente). Esta informação é pertinente para o desenvolvimento da indústria de duas rodas, designadamente localizada na região de Aveiro e Águeda.

4.4 Análise sobre as Tabelas de Contingência

4.4.1 Identificação da Escola

4.4.1.1 Utilização do Automóvel na deslocação casa/escola

Foram analisadas as respostas de acordo com duas escolas secundárias frequentadas pelos inquiridos (estudantes), a Escola Secundária Marques de Castilho e a Escola Secundária Adolfo Portela. Relativamente ao uso do automóvel na deslocação casa/escola, a maioria dos alunos da escola secundária Marques de Castilho usa o automóvel como meio de transporte no seu percurso casa/escola enquanto grande parte dos alunos da escola secundária Adolfo Portela não usam este meio de transporte no seu percurso casa/escola.

4.4.1.2 Perceção de barreiras na sua deslocação casa/escola, em bicicleta convencional

Os alunos que frequentam a Escola Secundária Adolfo Portela valorizam mais as barreiras “Risco de Acidente com outros veículos” e “Ausência de Conforto” do que os alunos da Escola Secundária Marques de Castilho. Por outro lado, estes últimos consideram os obstáculos “Ruído de Tráfego”, “Tráfego motorizado na infraestrutura (VPM)”, “Relevo”, “Estacionamento da Bicicleta”, “Pavimento Irregular”, “Condições Climáticas” e “Risco de Acidente com peões” aqueles que são de maior importância.

4.4.1.3 Regime de utilização da bicicleta elétrica

Quanto ao regime de utilização da bicicleta elétrica, foi observado que os alunos da escola secundária Adolfo Portela preferem comprar a bicicleta elétrica enquanto os alunos da escola secundária Marques de Castilho preferem pagar um aluguer por elas.

Identificação da Escola * Regime de Utilização da Bicicleta Elétrica Tabulação cruzada					
			Regime de Utilização da Bicicleta Elétrica		Total
			Compra da Bicicleta Elétrica	Aluguer mensal da Bicicleta Elétrica	
Identificação da Escola	ESAP - Escola Secundária Adolfo Portela	Contagem	15	6	21
		Contagem Esperada	9,3	11,7	21,0
		Resíduos ajustados	3,0	-3,0	
	ESMC - Escola Secundária Marques de Castilho	Contagem	17	34	51
		Contagem Esperada	22,7	28,3	51,0
		Resíduos ajustados	-3,0	3,0	
Total	Contagem		32	40	72
	Contagem Esperada		32,0	40,0	72,0

Tabela 4.6: Tabela de Contingência (Escola versus Regime de Utilização)

De acordo com a tabela de contingência os resíduos ajustados são bastante significativos, por esse motivo existe uma forte possibilidade de haver uma associação entre as duas categorias. A forte associação diz-nos que os alunos da Escola Secundária Adolfo Portela têm uma propensão maior de comprar a bicicleta elétrica do que os alunos da Escola Secundária Marques de Castilho, enquanto os alunos da Escola Secundária Marques de Castilho têm uma tendência mais significativa de optar pelo aluguer mensal da bicicleta elétrica. Contudo, é necessário efectuar o teste de independência com a finalidade de garantir a significância patente sobre essa associação. Com efeito, já havia sido verificado que os alunos da escola secundária Adolfo Portela usam menos o automóvel do que os alunos da escola secundária Marques de Castilho.

Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)	Sig exata (2 lados)	Sig exata (1 lado)
Qui-quadrado de Pearson	8,743 ^a	1	,003	,004	,003
Correção de continuidade ^b	7,268	1	,007		
Razão de verossimilhança	8,871	1	,003		
Teste Exato de Fisher					
Associação Linear por Linear	8,621	1	,003		
N de Casos Válidos	72				

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 9,33.

b. Computado apenas para uma tabela 2x2

Tabela 4.7: Teste de Independência Qui-quadrado e Exato de Fisher

Uma vez que não existe nenhuma célula na tabela de contingência com uma frequência inferior a 5, foi utilizada a estatística de teste qui quadrado com o fim de se saber se a associação existente é estatisticamente significativa ou não. Através do teste qui-quadrado é confirmada a forte associação existente entre os grupos descritos na tabela anterior. De acordo com o teste, é possível afirmar a existência de uma associação entre os grupos para qualquer nível de significância usual, isto é, rejeita-se a hipótese nula (hipótese de independência dos grupos) para qualquer nível usual de significância.

4.4.2 Género

4.4.2.1 Utilização do transporte escolar, automóvel e o percurso a pé na deslocação casa/escola

Foi analisada a variável género nas perguntas relativas ao meio de transporte utilizado na deslocação casa/escola, à perceção de barreiras no uso da bicicleta convencional, aos modelos de bicicleta elétrica e aos equipamentos adicionais de sistemas de tecnologia avançada na bicicleta elétrica. De acordo com a pergunta relativa ao meio de transporte utilizado no trajeto casa/escola, o grupo do género feminino prefere andar a pé ou mesmo usar o transporte escolar (autocarro). Quanto ao uso do automóvel no trajeto casa/escola, os estudantes do género masculino tem uma maior propensão de usar o automóvel do que as raparigas.

4.4.2.2 Perceção de barreiras na sua deslocação casa/escola, em bicicleta convencional

Relativamente à perceção de barreiras no uso da bicicleta convencional, surgiu apenas uma associação significativa, para qualquer nível usual de significância, ligada à barreira correspondente ao tráfego motorizado na infraestrutura (VLP). De acordo com o estudo

efectuado, o grupo de alunos do género feminino valoriza mais esta barreira do que o grupo de alunos do género masculino. Quanto à possibilidade de se nomearem outras barreiras apenas 7 alunos responderam à questão, sendo que o grupo de alunos do género masculino nomeiam a distância e a autorização dos pais como barreiras enquanto o grupo dos alunos do género feminino nomeiam a pista *downhill* pouco cuidada como barreira.

4.4.2.3 Características da bicicleta elétrica consideradas importantes

Enquanto o grupo do género masculino valoriza mais, como característica importante na bicicleta elétrica, o conforto da suspensão, o grupo do género feminino prefere indicar o esforço físico reduzido e a segurança pessoal como os atributos mais relevantes da bicicleta elétrica.

4.4.2.4 Modelos da bicicleta elétrica

De entre os modelos de bicicleta elétrica alternativos, os estudantes do grupo feminino preferem o modelo utilitário, já os estudantes do grupo masculino optaram pelo modelo montanha/desporto ou mesmo o modelo design personalizado.

4.4.2.5 Equipamento adicional: Sistema Cooperativo de Bicicletas

Relativamente aos equipamentos adicionais, apenas no sistema cooperativo de bicicletas foi obtida uma associação. Os estudantes do género masculino preferem mais este sistema do que os estudantes do género feminino. Surgiu também a opção de se considerarem todos os 4 equipamentos adicionais em simultâneo, para este caso o género que mais valorizou a alternativa anterior foi o género masculino.

4.4.3 Grupo socioeconómico

Foi feita uma análise à variável relativa à profissão do encarregado de educação, para tal foi preciso agrupá-las de acordo com a escala de Graffar. Esta última tem como principal critério de agrupamento o nível de instrução e a origem do rendimento familiar. Com base nestes critérios as profissões obtidas no caso estudo presente foram agrupadas em 3 conjuntos da seguinte forma:

- Grupo 1: Grupo Socioeconómico de elevado rendimento → Bacharelato; Licenciatura; Mestrado ou Doutoramento.
- Grupo 2: Grupo Socioeconómico de médio rendimento → 9ºano - 12ºano; nove ou mais anos de escolaridade.
- Grupo 3: Grupo Socioeconómico de baixo rendimento → 9ºano; até nove anos de escolaridade.

4.4.3.1 Regime de utilização da bicicleta elétrica

Relativamente ao regime da bicicleta elétrica, o grupo socioeconómico de baixo rendimento tem uma maior propensão em usar a bicicleta convencional do que os restantes grupos que preferem não optar por nenhum regime.

4.4.3.2 Modelos da bicicleta elétrica

O modelo Montanha/Desporto é o mais votado para o grupo socioeconómico de baixo rendimento, enquanto o modelo utilitário é o mais apreciado pelo grupo socioeconómico de elevado rendimento.

Capítulo 5

Modelação

5.1 Introdução

Serão estimados neste capítulo diversos modelos de regressão logística no âmbito do estudo de preferências relativamente à futura utilização da bicicleta elétrica através de 9 opções de mobilidade sequenciais representados no inquérito do LNEC (modelo de preferências declaradas), seguido de uma breve interpretação dos resultados provenientes dos mesmos. Através do inquérito de preferências declaradas foram recolhidas 2232 observações pois o mesmo inquirido teve que responder a 9 opções de mobilidade e desta forma, por cada inquirido foram obtidas 9 observações. Em cada uma das 9 opções de mobilidade são dadas a escolher apenas duas alternativas ao aluno: considerar manter o automóvel ou substituí-lo pela bicicleta elétrica nas viagens a curta distância. Cada uma das opções de mobilidade é caracterizada por 4 atributos: custo do trajeto casa/escola em cada modo de transporte, existência de ciclovias no percurso casa/escola, inserção na bicicleta elétrica de novas tecnologias de informação e comunicação - Novas TIC e tempo do percurso casa/escola relativamente ao automóvel (Alternativa 1) e à bicicleta elétrica (Alternativa 2). De acordo com a estrutura das opções de mobilidade, o presente caso estudo tem como principal objetivo determinar quais os atributos considerados importantes na valorização e na escolha da bicicleta elétrica.

Custo do percurso casa/escola

No caso do automóvel, o custo foi estimado de acordo com o gasto médio de combustível nas deslocações diárias casa/escola, numa distância de 5km [Plano de Mobilidade e Transportes da Região de Aveiro]. Por outro lado, o custo do percurso associado à bicicleta elétrica teve em conta o cálculo do consumo médio de uma bateria, no mesmo período, sendo tomado como um aluguer do veículo elétrico de duas rodas. De acordo com as 9 opções de mobilidade, este atributo apresenta três diferentes situações de escolha (níveis do atributo): o valor do aluguer da bicicleta elétrica ser equivalente ao valor do passe mensal atual, o valor do aluguer da bicicleta elétrica ser 25% ou 50% superior ao valor do aluguer mensal atual.

Existência de pista ciclável no trajeto casa/escola

A existência de uma pista ciclável apenas faz sentido considerar no caso do uso da bicicleta elétrica, sendo que na equação do modelo associado ao automóvel vai tomar o valor 0. O referido atributo, conforme a opção, foi subdividido em 3 alternativas; não disponibilizando cobertura de ciclovias, disponibilizando 50% ou 100% do trajeto habitual com cobertura de ciclovias.

TIC

A inserção das novas tecnologias de informação e comunicação é um atributo considerado apenas para o uso da bicicleta elétrica, sendo que o conjunto das escolhas para esta variável é binária (ou se inserem as novas tecnologias ou não).

Tempo de viagem

A duração da deslocação casa/escola, com o uso do automóvel, foi calculada em 20 minutos. De acordo com as 9 opções de mobilidade, são apresentadas três alternativas para este atributo tendo em conta o uso da bicicleta elétrica; 10, 20 ou 30 minutos adicionais de viagem.

5.2 Opções de Mobilidade e Codificação das variáveis

Opções de Mobilidade	Custo	Comprimento da Ciclovía	Novas TIC	Tempo de Viagem
Opção 1	Custo do passe mensal = aluguer da bicicleta elétrica	Sem cobertura de pista ciclável	Ausência das novas TIC	Tempo de viagem 10 minutos superior ao do automóvel
Opção 2	Aluguer da bicicleta 25% superior ao passe mensal	Existência de 50% de pista ciclável	Ausência das novas TIC	Tempo de viagem 10 minutos superior ao do automóvel
Opção 3	Aluguer da bicicleta 50% superior ao passe mensal	Existência de 100% de pista ciclável	Presença das novas TIC	Tempo de viagem 10 minutos superior ao do automóvel
Opção 4	Custo do passe mensal = aluguer da bicicleta elétrica	Existência de 50% de pista ciclável	Presença das novas TIC	Tempo de viagem 20 minutos superior ao do automóvel
Opção 5	Aluguer da bicicleta 25% superior ao passe mensal	Existência de 100% de pista ciclável	Ausência das novas TIC	Tempo de viagem 20 minutos superior ao do automóvel
Opção 6	Aluguer da bicicleta 50% superior ao passe mensal	Sem cobertura de pista ciclável	Ausência das novas TIC	Tempo de viagem 20 minutos superior ao do automóvel
Opção 7	Custo do passe mensal = aluguer da bicicleta elétrica	Existência de 100% de pista ciclável	Ausência das novas TIC	Tempo de viagem 30 minutos superior ao do automóvel
Opção 8	Aluguer da bicicleta 25% superior ao passe mensal	Sem cobertura de pista ciclável	Presença das novas TIC	Tempo de viagem 30 minutos superior ao do automóvel
Opção 9	Aluguer da bicicleta 50% superior ao passe mensal	Existência de 50% de pista ciclável	Ausência das novas TIC	Tempo de viagem 30 minutos superior ao do automóvel

Tabela 5.1: Opções de Mobilidade

Nome da variável	Codificação	Frequência de indivíduos	Frequência de observações	Nº de indivíduos que escolhem a bicicleta elétrica
Escolha	0 - escolhe o automóvel	115	763	Não se aplica
	1 - escolhe a bicicleta elétrica	96	292	
Diferença de Custo entre a bicicleta elétrica e o automóvel	-1.92 (0.28-2.2) - Aluguer da bicicleta igual ao passe mensal	121	354	85
	-1.82 (0.38-2.2) - Aluguer da bicicleta 25% superior ao passe mensal actual	119	352	49
	-1.75 (0.45-2.2) - Aluguer da bicicleta 50% superior ao passe mensal actual	119	349	51
Comprimento da Ciclovia	0 - Sem cobertura de pista ciclável	121	350	59
	0.5 - 50% de pista ciclável	119	352	73
	1 - 100% de pista ciclável	120	353	70
Novas TIC	0 - Ausência das novas TIC	121	702	81
	1 - Presença das novas TIC	120	353	82
Diferença do Tempo de viagem da bicicleta elétrica e do automóvel	10 (30-20) - Tempo de viagem 10 min superior ao do automóvel	122	354	74
	20 (40-20) - Tempo de viagem 20 min superior ao do automóvel	119	353	70
	30 (50-20) - Tempo de viagem 30 min superior ao do automóvel	117	348	49
Escola	1 - ES. Adolfo Portela	36	305	28
	2 - ES. Marques de Castilho	86	750	68
Género	0 - Género feminino	52	441	42
	1 - Género masculino	70	614	54
Idade	1 - < 18 anos	100	873	81
	2 - ≥ 18 anos	22	182	15
Experiência Anterior	1 - < 10 anos	46	390	31
	2 - ≥ 10 anos	76	665	65
Grupo socioeconómico	1 - Grupo de alto rendimento	19	147	17
	2 - Grupo de médio rendimento	65	579	50
	3 - Grupo de baixo rendimento	38	329	29

Tabela 5.2: Categorização das variáveis em estudo e a frequência de indivíduos por categoria

5.3 Modelo base

Após o enquadramento teórico do modelo de regressão logística [3.3], vamos aplicá-lo ao conjunto de dados do presente caso estudo.

5.3.1 Estratégias de Modelação

Com efeito, é preciso ter em conta as variáveis (atributos) consideradas relevantes de acordo com o que se pretende analisar no caso em estudo. Os objetivos do mesmo tem por base analisar qual o papel futuro da mobilidade elétrica nas deslocações casa/escola e identificar as características da bicicleta elétrica mais valorizadas pelas famílias. Como modelo base foi considerado um modelo apenas com os 4 atributos principais que formam as 9 opções de mobilidade, já anteriormente descritos.

A modelação dos dados foi efetuada com o software R, um software gratuito, recorrendo à biblioteca *nnet* para a estimação dos modelos. No modelo que se segue são consideradas 4 variáveis independentes; a diferença entre o custo da bicicleta elétrica e o custo do automóvel, a presença ou ausência da pista ciclável, a inserção ou não das novas tecnologias TIC e a diferença entre os tempos de viagem da bicicleta elétrica e do automóvel, sendo o principal objetivo comparar os atributos entre si e analisar quais aqueles que mais contribuem na escolha na substituição do automóvel pela bicicleta elétrica. Embora, inicialmente, se tenha utilizado a formulação geral do modelo multinomial *logit* (frequentemente usado em casos onde haja mais do que duas alternativas), posteriormente optou-se pela especificação do modelo de regressão logística, uma vez que a variável dependente, no presente caso, apresenta apenas duas categorias (0 - Manter o uso do automóvel e 1 - Substituir o automóvel pela bicicleta elétrica). Desta forma, o modelo multinomial equivalente consiste na apresentação de apenas duas alternativas de escolha. Por este motivo, é equivalente considerar um modelo multinomial com apenas duas alternativas ou considerar um modelo linear generalizado, com base na família binomial *logit*, onde a variável dependente é uma variável binária (modelo de regressão logística).

Numa análise de regressão logística é importante, tal como em qualquer técnica de construção de um modelo de análise estatística, encontrar o melhor modelo e o mais parcimonioso contando que este descreva razoavelmente a relação entre a variável resposta (variável dependente) e o conjunto de variáveis regressoras (variáveis independentes). Estas últimas também são denominadas covariáveis.

5.3.2 Qualidade de Ajustamento do modelo

Foram estimados dois modelos base, com variável dependente binária e considerando apenas as 4 variáveis independentes já mencionadas (Custo, Comprimento da pista ciclável, Inserção das novas TIC e o Tempo de viagem). Ambos os modelos têm subjacentes duas funções de utilidade, sendo a primeira representativa da utilidade do automóvel e a segunda da bicicleta elétrica. Uma vez que a alternativa associada ao automóvel será a classe de referência do modelo (apenas do modelo com constante), na sua função de utilidade está contida uma constante “C”, representando a disponibilidade do automóvel.

Nome do Modelo	Função de Utilidade
Modelo base com constante	$U_{1n} = C + \beta_1 \text{Custo}_A + \beta_2 \text{Comp}_A + \beta_3 \text{Tic}_A + \beta_4 \text{Time}_A$
	$U_{2n} = \beta_1 \text{Custo}_B + \beta_2 \text{Comp}_B + \beta_3 \text{Tic}_B + \beta_4 \text{Time}_B$
Modelo base sem constante	$U_{1n} = \beta_1 \text{Custo}_A + \beta_2 \text{Comp}_A + \beta_3 \text{Tic}_A + \beta_4 \text{Time}_A$
	$U_{2n} = \beta_1 \text{Custo}_B + \beta_2 \text{Comp}_B + \beta_3 \text{Tic}_B + \beta_4 \text{Time}_B$

Tabela 5.3: Funções de Utilidade

Cada uma das funções de utilidade está associada ao indivíduo n e à alternativa i (U_{in}).

Foram utilizados quatro critérios de adequabilidade do modelo de acordo com o conjunto de dados; o *AIC* (3.2.5), o *Log-Likelihood* (3.2.3), o Pseudo-Coeficiente de Determinação de *McFadden* (3.3.4.3) e o teste de razão de verosimilhanças (3.2.4.1).

Critério	Fórmula	Interpretação
<i>AIC</i>	$AIC = Desvio(modelo) + 2p, \text{ onde o desvio é dado por}$ $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n 2\omega_i \{ [y_i(\hat{\theta}_{s_i} - \hat{\theta}_{c_i}) - a(\hat{\theta}_{s_i}) + a(\hat{\theta}_{c_i})] \}$	Quanto menor for o seu valor, melhor é o ajustamento do modelo. Usualmente para comparação de modelos não encaixados.
<i>Log-likelihood</i>	$\sum_{i=1}^n l_i(\theta_i, \omega_i, \phi) = \sum_{i=1}^n \frac{\omega_i}{\phi} (y_i \theta_i - a(\theta_i)) + \sum_{i=1}^n c(y_i, \phi)$	Quanto maior for o seu valor, melhor será o ajustamento do modelo ao conjunto dos dados.
Pseudo-Coeficiente de Determinação de <i>McFadden</i>	$R_{McFadden}^2 = 1 - \frac{\log(L(M_N))}{\log(L(M_S))}$	Varia entre 0 e 1, quanto mais perto de 1 melhor o seu ajustamento ao conjunto dos dados.
Teste de razão de verosimilhanças	$\Lambda = -2\{l(\hat{\beta}_0) - l(\hat{\beta})\}$	Usualmente para comparação de modelos encaixados.

Tabela 5.4: Critérios de Ajustamento

5.3.3 Análise do modelo

Antes de se ajustar o modelo base, foram excluídos todos os *missing values* do conjunto de dados, isto é, todas as observações que apresentassem pelo menos uma característica omissa. Desta forma, a amostra, que no início apresentava 248 inquiridos e 2232 observações, reteve 122 inquiridos e 1055 observações para o estudo do modelo.

Características da amostra obtida no Inquérito	Características da amostra obtida na Modelação
248 Inquiridos	122 Inquiridos
2232 Observações	1055 Observações

Tabela 5.5: Características Amostrais

De acordo com as 9 opções de mobilidade apresentados no inquérito, foram definidos dois modelos base, um com a presença de uma constante representativa da disponibilidade do automóvel (Modelo base I), que será a classe de referência do modelo, e outro sem a mesma (Modelo base II). Ambos os modelos apresentam funções de utilidade observada (componente determinística) caracterizadas pelos 4 atributos (Custo, Comprimento de uma pista ciclável, Inserção de Novas TIC e Tempo de viagem) apresentados nas opções de mobilidade. Uma vez

que os modelos são encaixados, a sua comparação pode ser efetuada com base no teste de razão de verosimilhanças.

Output:

```
> anova(modelobase2,modelobase1, test = "Chisq")
Analysis of Deviance Table

Model 1: choice ~ Custo + CompPC + TIC + Time - 1
Model 2: choice ~ Custo + CompPC + TIC + Time
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      1051      1215.2
2      1050      1157.4  1    57.788 2.919e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De acordo com o teste de razão de verosimilhanças existe uma diferença estatisticamente significativa entre os dois modelos (*valor* $p \approx 0$), sendo selecionado o modelo base I. Com efeito, as medidas de qualidade de ambos os modelos são as seguintes.

	<i>Log-likelihood</i>	Coeficiente de Determinação de <i>McFadden</i>
Modelo base I	-578.6974	0.07011
Modelo base II	-607.5915	0.02367

Tabela 5.6: Medidas de Qualidade

De acordo com os dois critérios acima descritos, o modelo base I apresenta um melhor ajustamento aos dados do que o modelo base II. Sendo preferível reter o modelo base I no estudo.

Modelo base I	Parâmetro	Estimativa	Estatística de Teste	Valor p
Intercept (Disponibilidade do Automóvel)	C	-14.61446	-7.360	0.000
Custo	β_1	-7.65243	-7.141	0.000
Comprimento da Ciclovía	β_2	0.52572	2.920	0.003
Novas TIC	β_3	0.48605	3.258	0.001
Tempo de viagem	β_4	-0.04486	-4.963	0.000

Tabela 5.7: Sumário do modelo base I

A tabela 5.7 fornece as estimativas dos parâmetros do modelo base I. Procedendo a uma análise sobre as variáveis verificamos que todas elas são significativas para o modelo, isto é, qualquer um dos atributos envolvidos no modelo é considerado um factor influente na escolha do automóvel ou da bicicleta elétrica.

De acordo com as estimativas dos parâmetros, a disponibilidade do automóvel, o custo e o tempo de viagem produzem uma redução na probabilidade de se escolher a bicicleta elétrica como meio de deslocação casa/escola. Relativamente à estimativa do parâmetro associado à disponibilidade do automóvel, o seu sinal negativo indica que, havendo o automóvel como opção de transporte, os inquiridos preferem este último. Contudo, o comprimento da pista ciclável e a presença das novas TIC provocam um aumento na probabilidade da escolha da bicicleta elétrica.

Neste modelo as correlações entre os parâmetros são bastante reduzidas, satisfazendo as condições exigidas de acordo com os pressupostos do modelo de regressão logística.

Modelo base II	Parâmetro	Estimativa	Estatística de Teste	Valor p
Custo	β_1	0.199641	1.795	0.072
Comprimento da Ciclovía	β_2	0.273116	1.595	0.111
Novas TIC	β_3	0.377206	2.592	0.009
Tempo de viagem	β_4	-0.043869	-5.043	0.000

Tabela 5.8: Sumário do modelo base II

De acordo com tabela 5.8, nem todas as variáveis são significativas para o modelo ao nível de significância usual. Ao nível de significância de 10%, as variáveis “Custo”, “Novas TIC” e “Tempo de viagem” são estatisticamente significativas para o modelo, sendo a variável “Comprimento da Ciclovía” muito pouco relevante para o modelo.

Para este caso, a correlação do Custo com o comprimento da ciclovía e com o tempo de viagem é muito elevada, 0.42 e 0.81 respetivamente [Anexo 1]. Esta situação num modelo de regressão logística não é a mais adequada dado que a discriminação do impacto das variáveis independentes entre si torna-se mais pobre. De acordo com os resultados apresentados anteriormente, o modelo base II não é o adequado para o estudo, uma conclusão que foi reforçada nas medidas de ajustamento. Por este motivo foi retido o modelo base I.

5.4. Modelo global

5.4.1 Escolha de covariáveis

Foram consideradas novas covariáveis, a partir do modelo base I, que pudessem ser estatisticamente significativas para o modelo. De acordo com o teste de razão de verosimilhanças foram considerados modelos simples ajustados para cada uma dessas novas covariáveis, com base no Método de *Hosmer e Lemeshow* [Hosmer, D. W. e Lemeshow, S. (2000)]. Primeiramente é enunciada a sua metodologia seguida da sua aplicação no projeto:

Método de *Hosmer e Lemeshow*

Quando temos um elevado número de covariáveis, o método *stepwise* pode não ser o mais adequado ao problema. Para o efeito, foi recomendado por *Hosmer e Lemeshow* um método de seleção de covariáveis constituído por 4 fases.

Passo 1. Análise de cada uma das covariáveis

Numa primeira fase da seleção são criados modelos simples, apenas contendo uma covariável, com vista a testar quais as variáveis relevantes para o modelo separadamente. Este processo depende do tipo de covariável, se esta for contínua constrói-se um modelo de regressão simples, caso esta seja discreta é possível criar uma tabela de contingência da variável resposta versus as k categorias da covariável ou ajustar um modelo de regressão simples.

Ao ajustar os modelos de regressão simples, a significância das covariáveis é baseada nos valores p do teste de *Wald* ou do teste de razão de verosimilhanças. Independentemente do tipo de covariáveis, são consideradas significativas aquelas que apresentarem um valor p inferior a 0.25. Este valor, aparentemente elevado, deverá permitir seleccionar covariáveis que pareçam pouco importantes num modelo de regressão simples, mas que se revelem mais importantes quando passamos a uma regressão múltipla. Para além de se considerarem as variáveis seleccionadas na escolha por regressão simples, também deverão ser incluídas todas as covariáveis que o investigador considere relevantes para o caso estudo.

Passo 2. Ajustamento do modelo múltiplo e sua análise

Após a seleção de cada uma das covariáveis através de modelos de regressão simples, é ajustado um modelo de regressão múltipla contendo todas as covariáveis consideradas no passo 1. Com o modelo múltiplo é avaliada novamente a importância de cada uma das covariáveis presentes de acordo com:

- 1- Teste de *Wald*: Com base neste teste, são obtidos os valores p para os coeficientes dos parâmetros correspondentes às diferentes covariáveis.
- 2- Cada um dos coeficientes estimados é comparado com o coeficiente do modelo que apenas contém aquela covariável.

As covariáveis que não produzirem uma melhoria no modelo devem ser excluídas, isto é, aquelas que apresentarem valores p superiores a 0.25 ou que apresentarem grandes alterações nos valores dos coeficientes estimados.

Passo 3. Reajustamento do modelo múltiplo

A seguir é reajustado o modelo múltiplo com todas as covariáveis selecionadas no passo 2. Sendo que este novo modelo deve ser comparado com o criado no passo 2 através do teste de razão de verossimilhanças e mais uma vez analisada a importância de cada uma das suas covariáveis. Este processo só termina quando todas as covariáveis relevantes estiverem incluídas no modelo múltiplo e todas aquelas que foram excluídas não serem estatisticamente significativas e/ou não serem importantes para a análise.

Neste passo devem ainda ser consideradas todas as covariáveis que foram excluídas no passo 1, por ordem decrescente de valor p. Este procedimento tem como principal objetivo identificar as possíveis covariáveis que isoladas não são significativas, mas que na presença de outras passam a ser.

Passo 4. Introdução de interações no modelo

A escolha das interações a serem adicionadas no modelo deve ter em conta o seu contexto no caso estudo e a sua relevância para o modelo. Tendo efectuado a escolha das interações pertinentes para o modelo é aplicado o método de seleção dos passos anteriores, tendo em atenção que devem ser excluídas as interações com valores p superiores a 0.10 (ou 0.05) e aquelas que apresentem grandes alterações nos valores dos coeficientes.

Na tabela seguinte estão representadas as estimativas para cada uma das covariáveis, o teste de *Wald* e da *Deviance* e respectivos valores p.

Covariável	Estimativa	Teste de Wald	Valor p (Wald)	<i>Deviance</i>	Valor p (<i>Deviance</i>)
Custo	-6.782	-6.691	0.00000	46.324	0.00000
Comp. Cicl.	0.3733	2.202	0.0276	4.8758	0.02724
Novas TIC	0.42114	2.951	0.00317	8.6094	0.003344
Tempo de Viagem	-0.03855	-4.467	0.00000	20.406	0.00000
Género 1	-0.2694	-1.943	0.0521	3.7629	0.0524
Escola 2	0.1979	1.277	0.202	1.6539	0.1984
Exp. Ant. Bic. 2	0.4396	2.974	0.00294	9.096	0.002562
Grupo Sócio Ec. 2	-0.71687	-2.996	0.00274	9.0294	0.002657
Grupo Sóc. Ec. 3	-0.77419				
Idade	-0.04587	-0.250	0.802	0.062857	0.802

Tabela 5.9: Modelos Simples ajustados para cada uma das covariáveis

De acordo com o método de *Hosmer e Lemeshow* são retidas apenas as covariáveis que apresentem um valor p inferior a 0.25. Assim sendo, as covariáveis mantidas após este critério foram as seguintes:

- Género,
- Escola,
- Experiência Anterior em Bicicleta,
- Grupo Sócioeconómico.

De seguida, foi ajustado um modelo múltiplo (Modelo 1) com todas as variáveis seleccionadas no passo anterior. Na tabela seguinte são apresentados novamente os testes de *Wald* e da *Deviance* e respectivos valores p com o fim de avaliar quais as covariáveis a serem mantidas no modelo múltiplo.

Covariável	Estimativa	Teste de Wald	Valor p (Wald)	<i>Deviance</i>	Valor p (<i>Deviance</i>)
Custo	-7.795	-7.187	0.0000	46.324	0.0000
Comp. Cicl.	0.532	2.924	0.0034	8.75	0.0031
Novas TIC	0.4971	3.292	0.0009	11.307	0.0007
Tempo de Viagem	-0.0456	-4.99	0.0000	21.261	0.0000
Género 1	-0.1621	-1.073	0.2834	0.951	0.3295
Escola 2	0.1998	1.217	0.2237	1.498	0.2209
Exp. Ant. Bic. 2	0.4282	2.681	0.0073	9.795	0.0017
Grupo Sóc. Ec. 2	-0.6656	-2.944	0.0032	9.179	0.0024
Grupo Sóc. Ec. 3	-0.7536				

Tabela 5.10: Modelo múltiplo com todas as covariáveis seleccionadas (Modelo 1)

De acordo com o valor p obtido com base no teste de Wald e no teste de *Deviance*, as covariáveis que apresentarem valores p inferiores a 0.25 são aquelas que deverão ser mantidas no modelo. Com efeito, as covariáveis seleccionadas são as seguintes:

- Escola,
- Experiência Anterior em Bicicleta,
- Grupo Socioeconómico.

Foi ajustado, de seguida, um novo modelo múltiplo (Modelo 2) com as covariáveis selecionadas na tabela anterior. Os resultados são apresentados na seguinte tabela:

Covariável	Estimativa	Teste de Wald	Valor p (Wald)	Deviance	Valor p (Deviance)
Custo	-7.7837	-7.183	0.0000	46.324	0.0000
Comp. Cicl.	0.5317	2.927	0.0034	8.75	0.0031
Novas TIC	0.4969	3.293	0.0009	11.217	0.0008
Tempo de Viagem	-0.0456	-4.997	0.0000	21.261	0.0000
Escola 2	0.1854	1.134	0.2566	1.301	0.254
Exp. Ant. Bic. 2	0.4691	3.025	0.0024	9.795	0.0017
Grupo Sóc. Ec. 2	-0.6936	-3.057	0.0022	9.269	0.0023
Grupo Sóc. Ec. 3	-0.7759				

Tabela 5.11: Modelo 2

De acordo com a tabela anterior a covariável “Escola” apresenta o valor limite de exclusão, contudo foi mantida no modelo. Os modelos 1 e 2 foram comparados através do teste de razão de verosimilhanças, sendo apresentado o seguinte output:

Analysis of Deviance Table

```

Model 1: choice ~ Custo + Tempo + Exp + SocEco + TIC + CompPC + GEN +
Escola
Model 2: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1046      1135.6
2      1047      1136.7 -1   -1.1479    0.284

```

Com base no valor p do teste de razão de verosimilhanças foi selecionado o modelo 2.

De acordo com o modelo 1 (modelo múltiplo com todas as covariáveis selecionadas) resultante do teste de razão de verosimilhanças, a variável “Género” não é significativa para o modelo e por esse motivo não foi considerada no modelo corrente. Por outro lado, era esperado, dada a análise exploratória efetuada no capítulo 4, que houvesse uma estimativa significativa na covariável género e também no cruzamento desta com o atributo TIC pois foi concluído que os inquiridos do género masculino possuem uma preferência superior à dos inquiridos do género feminino relativamente a esse atributo [Wardman et al 2007].

$$\text{Modelo 2: Escolha} \sim -15.213 - 7.783 * \text{Custo} - 0.046 * \text{Tempo} + 0.469 * \text{Exp} + 0.4969 * \text{TIC} - 0.693 * \text{SocEco2} - 0.775 * \text{SocEco3} + 0.531 * \text{CompPC} + 0.1854 * \text{Escola}$$

De acordo com o método de Hosmer e Lemeshow foram novamente introduzidas no modelo todas as covariáveis que haviam sido dispensadas na análise univariada, dado que estas poderiam vir a ser significativas para o modelo. Este passo não é apresentado pois nenhuma delas foi considerada estatisticamente significativa para o modelo. Assim, procedeu-se à consideração de interações no modelo.

5.5 Interações

5.5.1 Fatores de agrupamento

Em relação às interações, é interessante saber se a experiência anterior em bicicleta do inquirido influencia a escolha deste último para qualquer uma das opções de mobilidade apresentadas. Uma vez que se verificou que o factor socioeconómico influencia na escolha do

estilo de veículo a adoptar, é possível que a sua interacção com alguns dos 4 atributos das opções de mobilidade seja significativa pois trata-se, neste caso, da escolha da bicicleta elétrica e não da convencional. Para além disso, dado que as respostas foram efetuadas por alunos de duas escolas secundárias, a variável “Escola” pode apresentar aspectos importantes relativos à enorme dependência no uso do automóvel no percurso casa/escola, nomeadamente na elevada percentagem de alunos da escola Marques de Castilho que usam o automóvel como meio de transporte casa/escola e na preferência dos mesmos pelo aluguer da bicicleta elétrica sobre a sua compra.

Para o efeito, foram adicionadas as seguintes interacções:

- Escola vs Comprimento da Pista Ciclável
- Escola vs Tempo de viagem
- Escola vs TIC
- Grupo Sócioeconómico vs Custo
- Grupo Sócioeconómico vs TIC

Quando as interacções foram introduzidas, apesar de nenhuma delas apresentar uma frequência de respostas igual a 0, apenas a interacção entre a Escola e as Novas TIC foi considerada estatisticamente significativa (valor $p < 0.25$). Esta nova interacção mostra que existe uma distinção entre a preferência da inserção das novas TIC entre as escolas secundárias Adolfo Portela e Marques de Castilho. Com efeito, a explicação para a sua significância pode estar relacionada com o facto de que a escola secundária Adolfo Portela engloba, na sua maioria, cursos na área das tecnologias e desta forma os alunos da mesma manifestarem uma maior motivação sobre a inserção das novas TIC.

De seguida é apresentada a tabela com o teste de Wald e respectivo valor p para cada uma das covariáveis do modelo.

Covariável	Estimativa	Teste de Wald	Valor p (Wald)
Custo	-7.8263	-7.196	0.0000
Comp. Cicl.	0.5334	2.924	0.0035
Novas TIC	1.4403	2.395	0.0166
Tempo de Viagem	-0.0458	-5.000	0.0000
Escola 2	0.3396	1.872	0.0612
Exp. Ant. Bic. 2	0.4542	2.915	0.0035
Grupo Sóc. Ec. 2	-0.6992	-3.061	0.0022
Grupo Sóc. Ec. 3	-0.7796		
Novas TIC * Escola	-0.5441	-1.615	0.1062

Tabela 5.12: Modelo obtido com o método de Hosmer e Lemeshow (Modelo 3)

Com a introdução da interacção entre a Escola e as Novas TIC, a covariável Escola tornou-se mais significativa para o modelo ($p = 0.06$).

Desta forma, o modelo com a interacção (Modelo 3), de acordo com o método de Hosmer e Lemeshow, é dado por

$$\text{Modelo 3: } Escolha \sim -15.65 - 7.83 * Custo - 0.05 * Tempo + 0.45 * Exp + 1.44 * TIC - 0.69 * SocEco2 - 0.78 * SocEco3 + 0.53 * CompPC + 0.34 * Escola - 0.54 * TIC * Escola$$

Uma vez que os modelos 2 e 3 são encaixados, o critério de seleção mais adequado é o teste de razão de verosimilhanças.

Analysis of Deviance Table

Model 1: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
 Model 2: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
 + TIC * Escola

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	1047	1136.7			
2	1046	1134.2	1	2.5179	0.1126

De acordo com o valor p do teste de razão de verossimilhanças, a diferença entre os dois modelos não é estatisticamente significativa e por esse motivo foi selecionado o modelo 2. Com efeito, as medidas AIC, *log-likelihood* e do coeficiente de determinação de *McFadden* são

Nome do Modelo	AIC	Log-likelihood	Coeficiente de Determinação de <i>McFadden</i>
Modelo 2	1152.7	-568.37	0.0867
Modelo 3	1152.2	-567.11	0.0887

Tabela 5.13: Modelo 2 vs Modelo 3

Sendo que o AIC mede a quantidade de informação perdida do modelo, assim o modelo que apresentar um menor valor desta medida é aquele que é selecionado.

De acordo com a tabela 5.13, o novo modelo (Modelo 3) produziu um menor AIC do que o modelo sem interação (Modelo 2), contudo essa melhoria não é estatisticamente significativa de acordo com o teste de razão de verossimilhanças. Assim, foi selecionado o modelo 2.

5.5.2 Método *Stepwise*

Através do método *stepwise* é possível também construir um outro modelo alternativo ao obtido no passo anterior. Para o efeito, foi criado um modelo com todas as covariáveis e todas as interações consideradas no passo anterior.

Assim, tem-se o seguinte modelo inicial:

*Modelo 4: Escolha ~ Custo + CompPC + TIC + Tempo + GEN + Escola + Soc.Eco. + Exp + Idade + GEN * TIC + Escola * CompPC + Escola * TIC + Escola * Tempo + Soc.Econ.*Custo + Soc.Eco.*TIC*

Foi também adicionada a interação entre o género e as novas TIC pois, de acordo com a análise exploratória efetuada (Capítulo 4), os estudantes do género masculino valorizam mais o equipamento adicional relativo ao sistema cooperativo de bicicletas, enquanto os alunos do género feminino valorizam menos esse atributo em relação ao grupo do género masculino.

De acordo com o método *stepwise*, o novo modelo (Modelo 5) apresenta as seguintes estimativas das suas covariáveis:

Covariável	Estimativa	Teste de Wald	Valor p (Wald)
Custo	-7.786	-7.170	0.0000
Comp. Cicl.	0.5309	-0.917	0.3591
Novas TIC	1.4672	2.445	0.0144
Tempo de Viagem	-0.0462	-5.033	0.0000
Escola 2	0.0371	0.127	0.8986
Exp. Ant. Bic. 2	0.4727	3.040	0.0023
Grupo Sóc. Ec. 2	-0.7011	-3.061	0.0021

Grupo Sóc. Ec. 3	-0.7838		
Comp P.C. * Escola	0.6874	1.713	0.0866
Novas TIC * Escola	-0.5608	-1.666	0.0956

Tabela 5.14: Modelo obtido com o método *stepwise* (Modelo 5)

De acordo com a tabela acima, com a permanência da interação da Escola com o Comprimento da Pista Ciclável a significância das covariáveis Escola e Comprimento da Pista Ciclável deixaram de ser estatisticamente significativas para o modelo. Contudo e uma vez que os modelos 2 e 5 são encaixados, o critério de comparação a ser aplicado é o teste de razão de verossimilhanças.

Analysis of Deviance Table

```
Model 1: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
Model 2: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
+ CompPC * Escola + TIC * Escola
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1047      1136.7
2      1045      1131.3  2    5.4625  0.06514 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De acordo com o teste de razão de verossimilhanças o modelo 5 é o mais adequado para o estudo ao nível de significância a 10%, contudo este último modelo possui duas covariáveis não significativas para o modelo (Comprimento da ciclovia e a Escola). Com vista a obter mais informação sobre os dois modelos, foram calculadas as medidas AIC, da log-verossimilhança e do coeficiente de determinação de *McFadden* são dadas por

Nome do Modelo	AIC	Log-verossimilhança	Coeficiente de Determinação de <i>McFadden</i>
Modelo 2	1152.7	-568.37	0.0867
Modelo 5	1151.3	-565.64	0.0911

Tabela 5.15: Modelo 2 vs Modelo 5

Através do critério AIC conclui-se que o modelo 5 é o mais adequado para o estudo, ainda que seja apenas uma pequena diferença. A medida de log-verossimilhança também aumentou (melhorou) e a medida associada ao ajustamento do conjunto de dados ao modelo também produziu uma melhoria do modelo 2 para o modelo 5. Por este motivo, foi selecionado o modelo 5 para o estudo, apesar de este possuir duas covariáveis não significativas (Comprimento da ciclovia e a Escola).

Modelo final:

Modelo 5: Escolha ~ - 14.958 - 7.786 * *Custo* - 0.0462 * *Tempo* + 0.4727 * *Exp* + 1.4672 * *TIC* - 0.7011 * *SocEco2* - 0.7838 * *SocEco3* + 0.5309 * *CompPC* + 0.0371 * *Escola* + 0.6874 * *CompPC* * *Escola* - 0.5608 * *TIC* * *Escola*

5.6 Análise de resíduos

5.6.1 Diagnóstico

Uma vez ajustado o modelo final, é necessário avaliar a qualidade do mesmo através da análise de resíduos, nomeadamente a escolha da distribuição, da função de ligação, dos valores preditos e das observações mal ajustadas (capacidade de discriminação do modelo).

Numa primeira abordagem procedeu-se à representação gráfica dos resíduos de *Pearson* padronizados versus os índices das observações. No gráfico 5.1 é possível observar uma ausência de padrão no conjunto das observações, os resíduos encontram-se espalhados na horizontal em torno do zero embora a sua média seja positiva (0.00465), quando deveria ser nula, dada a grande disparidade entre a frequência de resíduos negativos e positivos, 763 e 292, respetivamente. Esta situação é explicada por ter surgido uma frequência maior para a alternativa de manter o automóvel ($Y=0$) do que substituir este pela bicicleta elétrica ($Y=1$). Um padrão nulo é consistente com um modelo adequado.

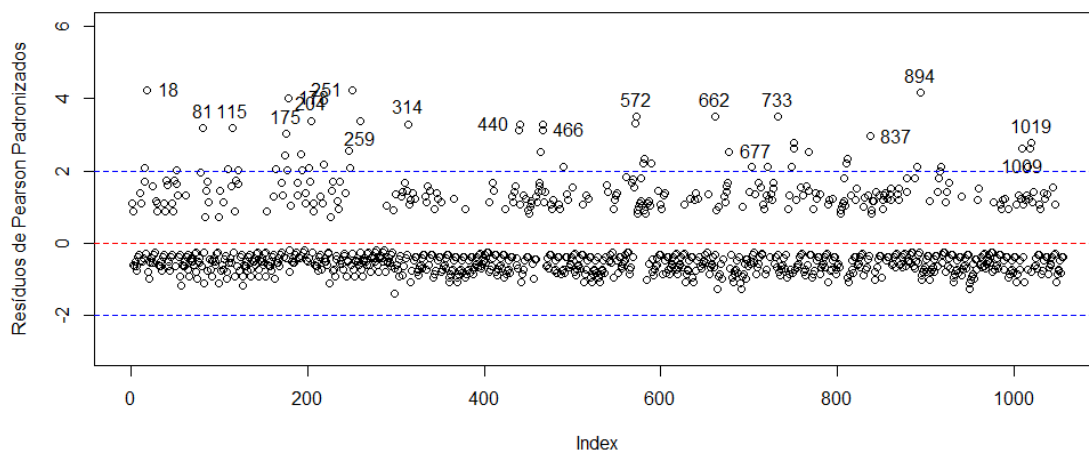


Figura 5.1: Resíduos *Pearson* padronizados vs Observações

Ainda é possível verificar no gráfico 5.1 a existência de algumas observações tomadas como *outliers*, perfazendo um total de 52 observações, sendo aquelas que se encontram assinaladas no gráfico as observações 18, 81, 115, 175, 178, 204, 251, 314, 440, 466, 572, 662, 677, 733, 837, 894, 1009 e 1019, por ordem de observação. Estes valores encontram-se afastados do intervalo considerado aceitável, -2 a 2, indicando um ajustamento mal efetuado sobre a variável dependente.

Modelo final:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -14.958 - 7.786 * Custo - 0.0462 * Tempo + 0.4727 * Exp + 1.4672 * TIC \\ - 0.7011 * SocEco2 - 0.7838 * SocEco3 + 0.5309 * CompPC + 0.0371 * Escola \\ + 0.6874 * CompPC * Escola - 0.5608 * TIC * Escola$$

De acordo com o modelo final, entre as 52 observações mal ajustadas, uma maioria apresenta características como, custo do aluguer da bicicleta elétrica 50% ou 25% superior ao passe mensal (a alternativa mais cara e média respetivamente), o tempo de viagem da bicicleta elétrica 30 ou 20 minutos superior ao do automóvel, o grupo socioeconómico de rendimento baixo ou médio, a ausência das novas TIC e a ausência de uma pista ciclável no trajeto casa/escola, variáveis que deveriam, em teoria, despromover a escolha da bicicleta elétrica, o que não foi o sucedido. Segundo o modelo final, os indivíduos com as características já referidas deveriam optar por manter o automóvel.

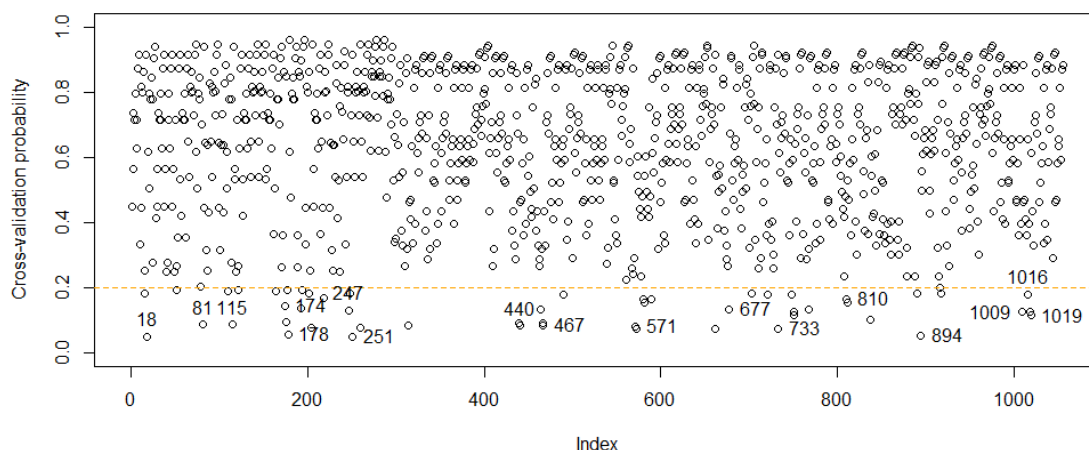


Figura 5.2: Probabilidades Cruzadas vs Observações

De acordo com a figura 5.2, verifica-se que a grande maioria das observações identificadas como *outliers* (gráfico 5.1), são aquelas que apresentam uma probabilidade reduzida de se realizarem. Sendo que os 52 *outliers* detectados no gráfico 5.1 são aqueles que produzem uma probabilidade inferior a 0.2 de se virem a concretizar.

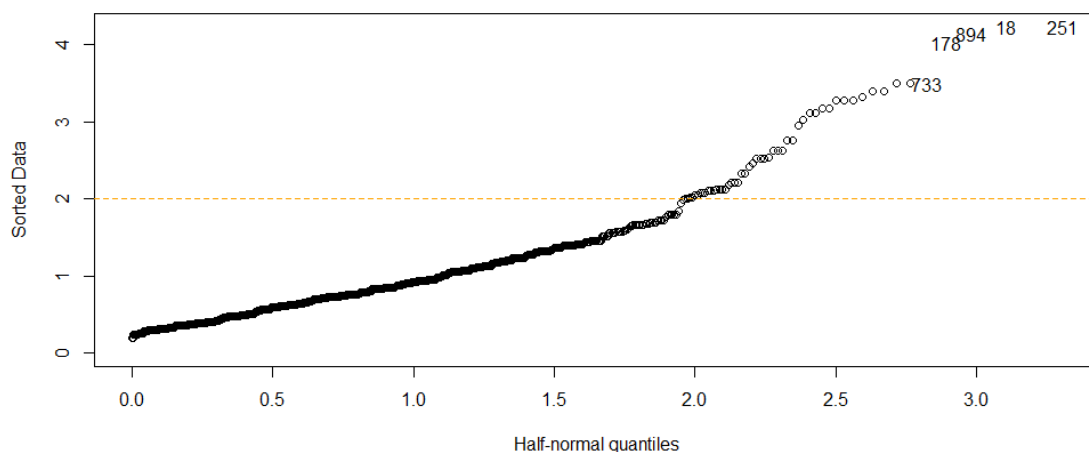


Figura 5.3: Halfnormal dos resíduos de *Pearson* padronizados

Na figura 5.3 é possível verificar que as observações 251, 18, 894, 178 e 733 poderão ser consideradas discordantes pois possuem valores extremais muito elevados e estes também já haviam sido tomados como observações detentoras de um valor residual acima do nível recomendado (observações invulgares).

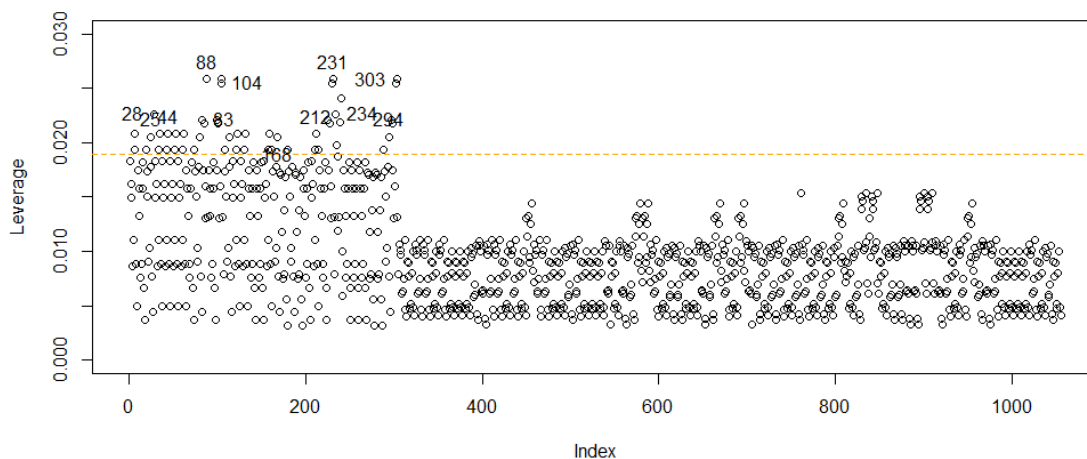


Figura 5.4: Leverages vs Observações

Quanto às observações consideradas influentes, de acordo com a figura 5.4, destacam-se as observações 25, 28, 44, 83, 88, 104, 168, 212, 231, 234, 294 e 303. Estas observações apresentam valores suficientemente elevados para porem em causa o modelo. A influência depende da resposta e das variáveis explicativas, sendo que as observações com *leverage* muito elevado, quando retiradas do conjunto dos dados, alteram o ajustamento do modelo. Por esta razão, deve ser estimado um novo modelo sem as observações referenciadas no gráfico 5.4 e que, em simultâneo estas produzam um valor residual muito afastado de zero (> 2), isto é, sejam também *outliers*. Foi obtido um total de 50 observações influentes.

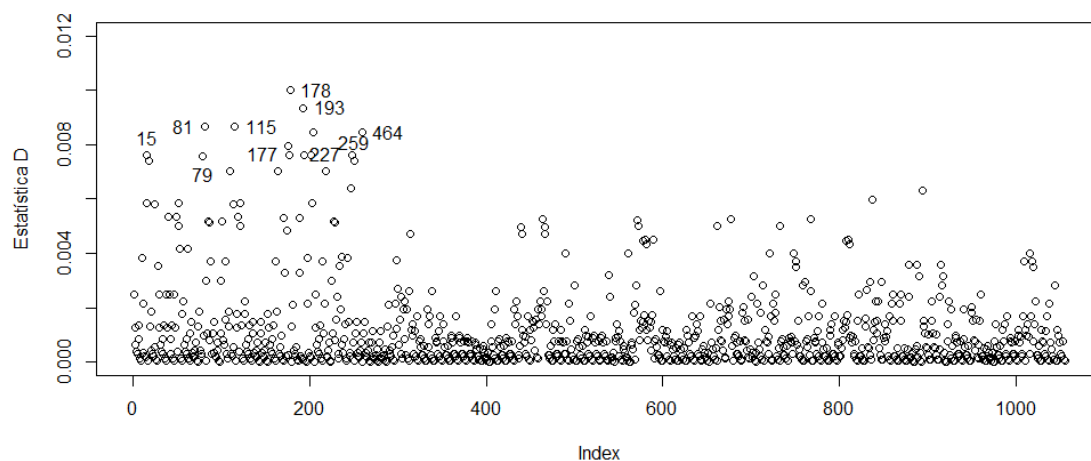


Figura 5.5: Estatística D vs Observações

Na figura 5.5 estão identificados os maiores valores de acordo com a estatística D. As observações que produziram um valor elevado desta estatística e que, em simultâneo, tiverem sido identificadas como *outliers* no gráfico 5.1, foram retiradas do conjunto de dados. De acordo com o critério, foram retirados do modelo 24 observações (15, 18, 81, 110, 115, 164, 175, 177, 178, 193, 194, 202, 204, 218, 247, 248, 251, 259, 464, 571, 677, 767, 837 e 894). De seguida, foi efectuado um reajustamento de um novo modelo com vista a determinar se a sua exclusão produzirá alterações significativas nas estimativas dos parâmetros ou levará mesmo à escolha de outras covariáveis.

5.6.2 Ajustamento do novo modelo

Foi reajustado um novo modelo com todas as covariáveis disponíveis, sem as 24 observações com valores elevados da estatística D e consideradas *outliers*. Sobre o modelo foi aplicado novamente o método *stepwise* e as estimativas, teste de *Wald* e respectivos valores p do novo modelo (Modelo 6) são os seguintes:

Covariável	Estimativa	Teste de Wald	Valor p (Wald)
Custo	-10.048	-8.248	0.0000
Comp. Cicl.	-0.411	-0.494	0.621
Novas TIC	1.942	2.872	0.0041
Tempo de Viagem	-0.066	-6.349	0.0000
Escola	0.451	1.376	0.1689
Exp. Ant. Bic.	0.707	4.185	0.0000
Grupo Sóc. Ec.	-0.481	-4.072	0.0000
Comp. Cicl. * Escola	0.686	1.503	0.1327
TIC * Escola	-0.795	-2.127	0.0334

Tabela 5.16: Modelo obtido pelo método *stepwise* sem as 24 observações (Modelo 6)

De seguida, foi efetuada uma comparação entre o modelo que havia sido selecionado antes da análise de resíduos e o agora obtido com base no critério AIC pois os modelos não são encaixados.

Modelos	AIC	Log-verossimilhança	Coefficiente de Determinação de McFadden
Modelo 5	1151.3	-565.64	0.0911
Modelo 6	1033.8	-506.88	0.1419

Tabela 5.17: Modelo 5 vs Modelo 6

De acordo com o critério AIC foi selecionado o modelo 6 uma vez que este apresenta uma menor quantidade de informação perdida (AIC). Para além disso, o coeficiente de determinação aumentou consideravelmente do modelo 3 para o modelo 6, indicando que este último apresenta um melhor ajustamento ao conjunto de dados.

Modelo Final (Modelo 6):

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -19.897 - 10.048 \text{Custo} - 0.411 \text{CompPC} + 1.942 \text{TIC} - 0.066 \text{Tempo} \\ + 0.451 \text{Escola} + 0.707 \text{Exp} - 0.481 \text{SocEco} + 0.686 \text{Comp} * \text{Escola} - 0.795 * \text{TIC} * \text{Escola}$$

Adequação da função de ligação

Após a análise de resíduos sobre o modelo, é necessário avaliar a adequação da função de ligação do mesmo. O teste visa verificar se o modelo possui uma função de ligação *logit* ou não, sendo que para este último caso terá de se identificar a verdadeira função de ligação. Irão ser apresentados dois testes, um baseado na inserção de uma nova covariável (z_i), sendo que esta representa a função de ligação *logit* para cada observação i , ao preditor linear e observar se há um declínio da função desvio e outro considerando $\hat{\eta}^2$ como uma nova covariável a adicionar ao preditor linear e verificar se produz o mesmo declínio na função desvio[Collett, D. (2003)].

Analysis of Deviance Table

```
Model 1: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
+ CompPC * Escola + TIC * Escola
Model 2: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
+ CompPC * Escola + TIC * Escola + zi
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1021      1013.8
2      1020      1013.1  1  0.60114  0.4381
```

De acordo com o teste de razão de verossimilhanças os modelos não possuem uma diferença estatisticamente significativa, pelo que não existem evidências para afirmar que a função de ligação *logit* não é a adequada.

Analysis of Deviance Table

```
Model 1: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
+ CompPC * Escola + TIC * Escola
Model 2: choice ~ Custo + Tempo + Exp + TIC + SocEco + CompPC + Escola
+ CompPC * Escola + TIC * Escola + eta2
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1021      1013.8
2      1020      1013.1  1  0.68754  0.407
```

De acordo com o teste de razão de verossimilhanças, adicionando a covariável $\hat{\eta}^2$, a diferença entre os modelos não é estatisticamente significativa, sendo que não existem evidências para afirmar a inadequabilidade da função de ligação *logit*.

5.7 Predição do modelo

Uma vez escolhido o modelo mais adequado de acordo com a sua qualidade de ajustamento ao conjunto dos dados, foi avaliada a capacidade de predição do mesmo, ou seja, a capacidade de discriminar as escolhas efetuadas por cada um dos indivíduos corretamente. No caso estudo presente, implica discriminar os indivíduos que optaram por manter o automóvel daqueles que preferiram substituir este pela bicicleta elétrica. Para o efeito aplicou-se o erro de predição e a curva ROC (tabela 5.19 e gráfico 5.8).

5.7.1 Erro de predição

Ajustados	Observados		
	0	1	Total
0	708 (68.7%)	204 (19.8%)	912
1	55 (5.3%)	64 (6.2%)	119
Total	763	268	1031

Tabela 5.18: Valores Ajustados (cutpoint = 0.5) vs Valores Observados

No cruzamento entre os valores observados e os valores ajustados superiores e inferiores a 0.5, para a variável dependente, foi obtida uma proporção total de casos corretamente preditos de cerca de 75%. As percentagens representadas entre parêntesis foram calculadas sobre o total das observações (1031).

5.7.2 Curva ROC

Foi ajustada uma curva ROC à capacidade de discriminação do modelo com uma maior precisão, visto que o *cutpoint* foi optimizado no sentido de se obter a melhor discriminação possível.

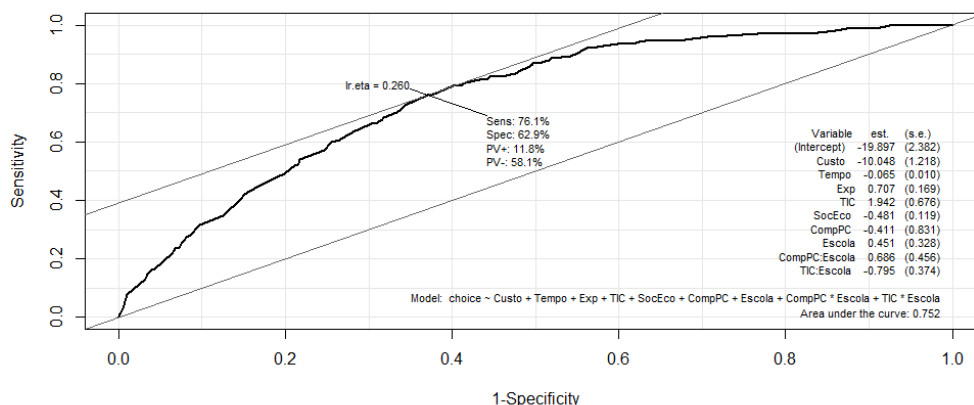


Figura 5.6: Curva ROC

Na figura 5.6, é possível observar uma análise de discriminação entre os valores preditos e observados do modelo tendo em conta duas medidas de avaliação para o efeito, a sensibilidade e a ausência de especificidade. O valor correspondente à área abaixo da curva ROC foi de 0.752, que de acordo com os níveis conhecidos indicam que o modelo detém uma capacidade de discriminação aceitável. Ainda é possível verificar que o valor relativo à sensibilidade (76.1%) é superior ao valor relativo à especificidade (62.9%), isto é, discrimina melhor os verdadeiros positivos ($Y=1$) do que os verdadeiros negativos ($Y=0$). Contudo, o nível de discriminação aceitável foi obtido com um *cutpoint* equivalente a 0.26, permitindo identificar correctamente uma maior percentagem de indivíduos que optaram pela bicicleta elétrica. No gráfico encontram-se também representadas as percentagens de valores preditos positivos e negativos mal ajustados, erro de tipo I e II respetivamente. Assim, o modelo contém cerca de 11.8% de respostas positivas ($Y=1$) e 58.1% de respostas negativas ajustadas incorrectamente.

Este resultado é o esperado tendo em conta a forte dependência no uso do automóvel e consequente reduzida adesão da bicicleta elétrica e barreiras que ainda há para solucionar. Para além disso, o mesmo mostra também que são exigidas medidas complementares por parte do município no futuro para fazer com que essa adesão aumente na deslocação casa/escola.

Capítulo 6

Interpretação dos resultados

6.1 Análise Exploratória

Neste capítulo irá proceder-se a uma análise estatística onde, numa primeira fase do inquérito, foram avaliados, fundamentalmente, o meio de transporte diário mais utilizado pelos estudantes, as barreiras sobre a deslocação da bicicleta e as características, sobre a mesma, consideradas mais relevantes. Na 2ª parte do inquérito foi avaliado, essencialmente, o tipo de regime de utilização de bicicleta elétrica adotado pelos inquiridos de acordo com o seu grupo socioeconómico. Na 3ª parte do inquérito os inquiridos (estudantes) foram

confrontados com várias opções de mobilidade (9), tendo que no fim escolher o automóvel ou a bicicleta elétrica.

Na 1ª parte do inquérito foram efetuadas questões, visando analisar o transporte mais usual nas deslocações casa/escola, o nível da adesão na mudança para a bicicleta elétrica, as barreiras mais valorizadas por parte dos estudantes, de modo a ser possível o seu uso no futuro, e as características mais importantes na bicicleta elétrica. Verificou-se que o automóvel é o transporte mais utilizado pelos estudantes (cerca de 50%), sendo um fator indicador de uma grande dependência no seu uso como era esperado. Para além disso, apenas uma pessoa usa a bicicleta convencional no percurso casa/escola (0.4%), revelando neste caso que é imperativo criar soluções motivadoras para o uso do transporte de duas rodas. Por outro lado, apesar de 99% dos inquiridos saber andar de bicicleta, apenas 16% afirmam andar regularmente de bicicleta convencional e só 6% (15 inquiridos) utilizou a bicicleta elétrica de Águeda. Na questão da valorização das barreiras no percurso casa/escola em bicicleta elétrica, as condições climatéricas e o pavimento irregular foram as consideradas mais importantes por parte dos inquiridos, cerca de 53% e 40%, respetivamente, sendo que apenas 15 inquiridos responderam a esta questão. Na perceção de barreiras no percurso casa/escola em bicicleta convencional foi obtida um total de 200 respostas, sendo que as barreiras de maior importância consideradas pelos inquiridos foram o risco de acidentes com outros veículos, o relevo e as condições climatéricas, com cerca de 33.5%, 30.5% e 25%, respetivamente. Através da questão relativa às barreiras existentes no município, foi possível identificar que as condições climatéricas é a barreira mais valorizada pelos inquiridos, independentemente do tipo de bicicleta (convencional ou elétrica), e que a barreira “relevo” deixa de existir quando se passa da bicicleta convencional para a bicicleta elétrica. Relativamente às características da bicicleta elétrica consideradas importantes por parte dos inquiridos, a acessibilidade no preço da bicicleta elétrica foi o atributo mais valorizado pelos inquiridos, cerca de 60%, sendo que houve cerca de 26% dos inquiridos que também valorizaram o esforço reduzido e a segurança rodoviária como características importantes. Este último resultado revela uma grande preocupação, por parte dos inquiridos, sobre as questões de segurança relativamente ao veículo de duas rodas e o seu preço de aquisição e manutenção.

Analisando por escolas, relativamente ao uso do automóvel no percurso casa/escola a maioria dos alunos da escola secundária Marques de Castilho usam o mesmo com maior frequência do que os alunos da escola secundária Adolfo Portela. Os alunos da escola secundária Adolfo Portela valorizam mais as barreiras “risco de acidente com outros veículos” e “ausência de conforto” do que os alunos da escola secundária Marques de Castilho.

Analisando por género, o grupo do género masculino tem uma maior propensão de usar o automóvel do que o grupo do género feminino. De acordo com a perceção de barreiras em bicicleta convencional, o tráfego motorizado na infraestrutura (VLP) é uma barreira mais valorizada pelo grupo do género feminino. Relativamente às características da bicicleta elétrica, o grupo do género masculino valoriza mais o conforto da suspensão enquanto o grupo do género feminino valoriza mais o esforço físico reduzido e a segurança pessoal.

Na 2ª parte do inquérito foram efetuadas questões relativas às preferências dos inquiridos, com vista a identificar o tipo de regime de utilização da bicicleta elétrica preferencial adoptado pelos mesmos, o modelo favorito a usar e o sistema de equipamento tecnológico a adicionar ao seu veículo elétrico de duas rodas. Relativamente ao regime de utilização da bicicleta elétrica, os inquiridos preferem optar pela utilização da bicicleta convencional (cerca de 50%) em alternativa à compra (cerca de 14%) ou aluguer (cerca de 17%) da bicicleta elétrica, revelando uma grande contenção sobre o custo de compra ou mesmo aluguer do veículo elétrico de duas rodas. O modelo preferido pela maioria dos inquiridos foi o modelo “Montanha/Desporto” (cerca de 52%), seguido do modelo “utilitário” (cerca de 26%) e do modelo “design personalizado” (cerca de 21%). Por outro lado, houve cerca de 51% dos inquiridos que afirmaram preferir adicionar novas TIC, sendo que 50% preferiu o sistema de TIC (GPS), 43% os 4 equipamentos em simultâneo e 25% o sistema de contabilização de calorias. Este resultado mostra uma grande necessidade de se desenvolver a indústria de

veículos de duas rodas para além do factor tecnológico servir de motivação na adesão ao veículo elétrico de duas rodas.

Analisando por escolas, relativamente ao regime de utilização da bicicleta elétrica, os alunos da escola secundária Adolfo Portela preferem comprar a bicicleta elétrica enquanto os alunos da escola secundária Marques de Castilho preferem alugar a mesma.

Analisando por género, entre os modelos da bicicleta elétrica, o grupo do género feminino prefere o modelo utilitário, enquanto o grupo do género masculino prefere o modelo “montanha/desporto” ou o modelo “design personalizado”. No caso dos equipamentos adicionais, o grupo do género masculino valoriza mais o sistema cooperativo de bicicletas e os 4 equipamentos adicionais em simultâneo do que o grupo do género feminino.

Analisando por grupo socioeconómico, o grupo socioeconómico de baixo rendimento prefere usar a bicicleta convencional, os grupos de médio e alto rendimento preferem não optar por nenhum regime. Relativamente aos modelos de bicicleta elétrica, o modelo “montanha/desporto” é mais valorizado pelo grupo socioeconómico de baixo rendimento, enquanto o modelo utilitário é mais valorizado pelo grupo socioeconómico de alto rendimento.

6.2 Modelo base obtido

Na 3ª parte do inquérito foram criados 9 opções de mobilidade sequenciais, cada um constituído por 4 atributos nomeadamente; a diferença de custo entre a bicicleta elétrica e o automóvel, a existência de uma pista ciclável ou não, a presença das novas tecnologias de informação e comunicação e a diferença da duração do tempo de viagem no trajeto casa/escola entre a bicicleta elétrica e o automóvel. Cada uma das 9 possibilidades de escolha visa avaliar a decisão do aluno em manter o automóvel ou substituir este último pela bicicleta elétrica no seu trajeto casa/escola. De seguida são apresentadas as diferentes características dos 4 atributos para cada uma das 9 opções de mobilidade e é efetuada uma análise sobre o modelo base I (modelo que contém apenas os 4 atributos referidos), através do *odds ratio*.

Para o modelo base I foi efetuada uma interpretação do *odds ratio* sobre cada uma das covariáveis presentes no modelo. Para o efeito, as covariáveis foram categorizadas e analisadas de acordo com o *odds ratio*.

Covariável	Estimativa	Estatística de Wald	Valor p	Odds Ratio	IC(95%)
Intercept	-0.298	-1.701	0.088	0.742	(0.524,1.045)
Custo 2	-1.062	-6.102	0.000	0.346	(0.245,0.484)
Custo 3	-1.181	-6.406	0.000	0.307	(0.212,0.438)
Comp. Cicl. 2	0.264	1.386	0.165	1.303	(0.896,1.898)
Comp. Cicl. 3	0.512	2.82	0.004	1.67	(1.173,2.395)
TIC 1	0.4104	2.594	0.009	1.507	(1.102,2.052)
Tempo de viagem 2	-0.394	-2.26	0.023	0.675	(0.478,0.947)
Tempo de viagem 3	-0.883	-4.797	0.000	0.414	(0.287,0.591)

Tabela 6.1: Sumário do modelo base

6.2.1 Interpretação das covariáveis

Custo

A diferença entre a classe de referência (Diferença de Custo da bicicleta elétrica com o do automóvel igual -1.92) e as restantes categorias (Custo 2 =-1.82 e Custo 3 =-1.75) é estatisticamente significativa, sendo os seus valores p próximos de 0.

De acordo com o *odds ratio*, a adesão da escolha da bicicleta elétrica diminui com a diminuição da diferença do custo da bicicleta elétrica com o custo do automóvel para -1.82, em valor absoluto, em cerca de 65%. Quando comparado com a diferença do custo da bicicleta elétrica com o custo do automóvel para -1.75, a adesão da escolha da bicicleta elétrica ainda é mais baixa que no caso anterior, tendo um decréscimo de cerca de 69%.

Em suma, quanto mais elevado e mais próximo do custo operacional do automóvel for o custo operacional da bicicleta elétrica menor será a adesão na escolha da mesma por parte dos inquiridos.

Comprimento da ciclovía

A diferença entre a classe de referência (inexistência de ciclovía no trajeto casa/escola) e a existência de 50% de uma ciclovía no trajeto casa/escola não é estatisticamente significativa para os níveis usuais (5% e 10%), por outro lado a diferença entre a classe de referência e a existência de 100% de uma ciclovía no trajeto casa/escola é estatisticamente significativa.

A adesão da escolha da bicicleta elétrica aumenta 1.3 e 1.7 vezes, respetivamente quando se passa a ter cerca de 50% e 100% de ciclovía no trajeto casa/escola. É possível verificar também que a adesão aumenta com o acréscimo da ciclovía na deslocação.

Novas TIC

Relativamente às novas tecnologias de informação e comunicação, a diferença entre a classe de referência (ausência das novas TIC) e a presença das novas TIC é estatisticamente significativa. O seu *odds ratio* indica que a presença das novas TIC nas bicicletas elétricas promove a sua escolha em 1.5 vezes, sendo este um indicador de muito interesse para o desenvolvimento da indústria do veículo de duas rodas.

Tempo de viagem

Para o caso do tempo de viagem na deslocação casa/escola, a diferença entre a classe de referência (tempo de viagem da bicicleta elétrica apenas 10 minutos superior ao tempo de viagem do automóvel) e as restantes classes são estatisticamente significativas.

Verifica-se que quanto maior for o tempo de viagem para a bicicleta elétrica, menor é a adesão na escolha da mesma, sendo o esperado. O decréscimo na adesão da escolha da bicicleta elétrica é em cerca de 33% e de 59% para as classes relativas ao tempo de viagem da bicicleta elétrica 20 e 30 minutos superior ao do automóvel, respetivamente.

6.2.2 Interpretação das Opções de mobilidade sequenciais

Como no presente caso estudo existem apenas duas alternativas (manter o automóvel ou substituir pela bicicleta elétrica), faz sentido avaliar a probabilidade de um inquirido vir a escolher a bicicleta elétrica de acordo com cada uma das 9 opções de mobilidade sequenciais e desta forma avaliar quais as opções de mobilidade mais favoráveis e os mais preteridos nesse sentido.

Perfis	Odds Ratio
Opção 1	0.69
Opção 2	0.42
Opção 3	0.52
Opção 4	0.93
Opção 5	0.35
Opção 6	0.12
Opção 7	0.48
Opção 8	0.21
Opção 9	0.09

Tabela 6.2: Perfis

De acordo com o odds ratio, é possível observar que qualquer uma das 9 opções de mobilidade, sendo que todos eles possuem um *odds ratio* entre 0 e 1, despromove a escolha da bicicleta elétrica. Contudo as opções de mobilidade que menos contribuem na escolha do automóvel e, por consequência, mais na escolha da bicicleta elétrica são, por ordem de importância, as opções 4, 1 e 3. Por outro lado, as opções que mais contribuem na adesão à escolha do automóvel e, por isso, menos na adesão à escolha da bicicleta elétrica são as opções 9, 6 e 8.

Opção 4	<ul style="list-style-type: none"> • Custo mais barato (0.28€), • Pelo menos a existência de 50% de pista ciclável no seu trajeto casa/escola, • Inserção das novas TIC, • Tempo de viagem 20 minutos superior ao do automóvel.
Opção 1	<ul style="list-style-type: none"> • Custo mais barato (0.28€), • Sem cobertura de pista ciclável, • Ausência das novas TIC, • Tempo de viagem 10 minutos superior ao do automóvel.
Opção 3	<ul style="list-style-type: none"> • Custo mais caro (0.45€), • Existência de pista ciclável em todo o seu trajeto, • Inserção das novas TIC, • Tempo de viagem 10 minutos superior ao do automóvel.

Tabela 6.3: Perfis mais votados

De acordo com a tabela 6.3, o atributo mais valorizado pelos inquiridos é o custo, revelando uma significativa diferença entre o consumo do automóvel e da bicicleta elétrica pois mesmo o custo mais caro da bicicleta elétrica (0.45€, apresentado na opção 3) continua a ser inferior ao consumo do automóvel (2.2€). Outro atributo considerado importante na escolha dos inquiridos é o tempo de viagem, apesar da opção mais votada ter sido aquela onde o tempo de viagem é 20 minutos superior ao tempo de viagem efectuado pelo automóvel.

Opção 9	<ul style="list-style-type: none"> • Custo mais caro (0.45€), • Pelo menos a existência de 50% de pista ciclável no seu trajeto casa/escola, • Ausência das novas TIC, • Tempo de viagem 30 minutos superior ao do automóvel.
Opção 6	<ul style="list-style-type: none"> • Custo mais barato (0.45€), • Sem cobertura de pista ciclável, • Ausência das novas TIC, • Tempo de viagem 20 minutos superior ao do automóvel.
Opção 8	<ul style="list-style-type: none"> • Custo mediano (0.38€), • Sem cobertura de pista ciclável, • Inserção das novas TIC, • Tempo de viagem 30 minutos superior ao do automóvel.

Tabela 6.4: Perfis menos votados

Em relação à tabela 6.4, o custo revela ser um factor importante na escolha dos inquiridos dado que os perfis com os custos mais caros da bicicleta elétrica foram os menos votados. Outros atributos também considerados importantes na escolha da bicicleta elétrica são o

tempo de viagem e a inserção das novas TIC (os perfis com os tempos de viagem mais longos foram os menos votados e dois dos três perfis menos votados apresentam a ausência das novas TIC), sendo este último atributo indicador de um interesse crescente, por parte dos inquiridos, sobre o avanço das novas tecnologias de informação e comunicação ligadas ao meio de transporte de duas rodas. Com efeito, no capítulo 4, em relação à questão do equipamento adicional, havia sido concluído que, entre os 4 equipamentos disponíveis, os inquiridos preferiram maioritariamente o sistema de tecnologia de informação e comunicação.

6.3 Modelo final obtido

Uma vez modelado o conjunto dos dados e tendo sido efetuada uma análise de qualidade sobre o modelo final ajustado, procedeu-se à interpretação das covariáveis do mesmo. Para o efeito, na tabela seguinte são apresentadas as estimativas de cada uma das covariáveis retidas no modelo, a estatística de *Wald* e seu respectivo valor p, assim como a estimativa do *odds ratio* e o correspondente IC estimado ao nível de confiança 95%. De acordo com o enquadramento teórico, descrito no capítulo 3, a interpretação dos modelos de regressão logística é obtida através da medida do *odds ratio* de cada uma das estimativas das covariáveis do modelo. Após a apresentação dos *odds ratio* de cada uma das covariáveis do modelo, é feita a interpretação das mesmas individualmente seguidas de uma discussão sobre o modelo e posteriormente as conclusões.

Covariável	Estimativa	Teste de Wald	Valor p	Odds Ratio	IC Odds Ratio
Intercept	-0.266	-0.724	0.469	0.767	(0.368,1.559)
Custo 2	-1.332	-6.753	0.000	0.264	(0.177,0.385)
Custo 3	-1.541	-7.214	0.000	0.214	(0.139,0.322)
Comp. Cicl. 2	0.071	0.166	0.868	1.073	(0.464,2.479)
Comp. Cicl. 3	0.227	0.553	0.58	1.255	(0.559,2.825)
Novas TIC 1	1.076	3.078	0.002	2.932	(1.473,5.826)
Tempo de Viagem 2	-0.615	-3.211	0.001	0.54	(0.369,0.783)
Tempo de Viagem 3	-1.302	-6.111	0.000	0.272	(0.177,0.409)
Escola 2	0.379	1.095	0.273	1.462	(0.749,2.935)
Exp. Ant. Bic. 2	0.699	4.106	0.000	2.012	(1.447,2.821)
Grupo Sóc. Ec. 2	-0.815	-3.78	0.000	0.443	(0.289,0.676)
Grupo Sóc. Ec. 3	-1.048	4.388	0.000	0.351	(0.219,0.559)
Novas TIC 1 * Escola 2	-0.889	-2.25	0.024	0.60	(0.188,0.889)
Comp. Cicl. 2 * Escola 2	0.58	1.192	0.233	2.61	(0.689,4.676)
Comp. Cicl. 3 * Escola 2	0.772	1.65	0.098	3.16	(0.866,5.456)

Tabela 6.5: Sumário do modelo final

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -0.266 - 1.332Custo2 - 1.541Custo3 + 0.076CompPC2 + 0.227CompPC3 \\ + 1.076TIC - 0.615Tempo2 - 1.302Tempo3 + 0.379Escola + 0.699Exp \\ - 0.815SocEc2 - 1.048SocEc3 - 0.889TIC1 * Escola + 0.58CompPC2 * Escola \\ + 0.772CompPC3 * Escola$$

6.3.1 Interpretação das covariáveis

Custo

As diferenças entre a classe de referência (Diferença de Custo da bicicleta elétrica com o do automóvel igual -1.92) e as restantes categorias (Custo 2 = -1.82 e Custo 3 = -1.75) são estatisticamente significativas, para qualquer nível de significância usual.

Com base no *odds ratio*, é possível observar que quanto menor for a diferença do custo da bicicleta elétrica com o do automóvel, isto é, quanto mais elevado for o custo da bicicleta menor é a adesão na escolha da mesma. O decréscimo que incorre nas classes Custo 2 e Custo 3 é de cerca de 74% e de 78%, respetivamente.

Comprimento da Ciclovía

As diferenças entre a classe de referência (inexistência de uma ciclovía) e as restantes classes não são estatisticamente significativas, apresentando valores p de 0.87 e 0.58 respetivamente. No entanto, é possível verificar um aumento na adesão da escolha da bicicleta elétrica à medida que a percentagem de pista ciclável existente no trajeto casa-escola vai aumentando em relação à classe de referência. Quando comparadas com a classe de referência (inexistência de ciclovía) ocorre um aumento na adesão da escolha da bicicleta elétrica de 1.1 e de 1.2 vezes, respetivamente.

Novas TIC

A diferença entre a classe de referência (ausência das novas TIC) e a presença das novas TIC é estatisticamente significativa (valor $p = 0.002$), para qualquer nível de significância usual. A presença das novas tecnologias de informação e comunicação provoca um aumento de 2.9 vezes no sentido da escolha da bicicleta elétrica quando comparada com a ausência da mesma.

Tempo de Viagem

Relativamente ao tempo de viagem, as diferenças entre a classe de referência (tempo de viagem da bicicleta elétrica apenas 10 minutos superior ao tempo de viagem do automóvel) e as restantes classes (tempo de viagem da bicicleta elétrica 20 e 30 minutos superior ao tempo de viagem do automóvel) são estatisticamente significativas, para qualquer nível de significância usual.

Foi obtido um decréscimo sobre a escolha da bicicleta elétrica, sendo essa descida de 46% e de 73% para o tempo de viagem da bicicleta elétrica 20 e 30 minutos superior ao do automóvel, respetivamente.

Escola

A diferença entre a classe de referência (Escola Secundária Adolfo Portela) e a Escola Secundária Marques de Castilho não é estatisticamente significativa, para qualquer nível de significância usual. De acordo com o *odds ratio*, os alunos da escola secundária Marques de Castilho possuem uma maior adesão na escolha da bicicleta elétrica do que os alunos da escola secundária Adolfo Portela, sendo que esse aumento de adesão é de 1.46 vezes. Este aumento pode ser justificado pela razoável diferença entre o número total de alunos, que responderam às opções de mobilidade sequenciais, da escola secundária Marques de Castilho (86) e da escola secundária Adolfo Portela (36), podendo assim fazer sentido haver mais alunos da escola secundária Marques de Castilho a optarem pela bicicleta elétrica do que os alunos da escola secundária Adolfo Portela.

Experiência Anterior em bicicleta

De acordo com a experiência anterior em bicicleta, a diferença entre a classe de referência (Experiência < 10) e a classe relativa à experiência ≥ 10 é estatisticamente significativa. Através do *odds ratio* é possível observar que os alunos com mais ou igual a 10 anos de experiência anterior em bicicleta convencional possuem uma maior propensão de vir a optar

pela bicicleta elétrica, sendo que essa propensão aumenta cerca de 2 vezes sobre os alunos com menos de 10 anos de experiência anterior em bicicleta convencional.

Grupo Socioeconómico

A diferença entre a classe de referência (grupo sócio económico de elevado rendimento) e as restantes classes (grupo sócio económico de médio e baixo rendimento) é estatisticamente significativo. De acordo com o *odds ratio*, quanto mais baixo for o rendimento do agregado familiar, menor é a adesão na escolha da bicicleta elétrica. Assim, os grupos sócio económicos de médio e baixo, quando comparados com o grupo socioeconómico de elevado rendimento, produzem um decréscimo, na escolha da bicicleta elétrica, de 56% e de 65%, respetivamente.

Se se tomar como classe de referência o grupo socioeconómico de médio rendimento, a diferença entre o grupo socioeconómico de baixo rendimento e a nova classe de referência não é estatisticamente significativa. Tendo sido obtido um valor p de cerca de 0.19. Por outro lado, a diferença entre a nova classe de referência e o grupo socioeconómico de alto rendimento é estatisticamente significativa, tendo produzido um valor p de cerca de 0.00015. Contudo, é possível observar através do *odds ratio* que, quando se passa do grupo socioeconómico de médio rendimento para o de baixo rendimento, ocorre um decréscimo na adesão da escolha da bicicleta elétrica em cerca de 21%, enquanto no caso da passagem do grupo socioeconómico de médio rendimento para o de alto rendimento a adesão na escolha da bicicleta elétrica aumenta cerca de 2.3 vezes.

Tomando o grupo socioeconómico de baixo rendimento como classe de referência, constata-se novamente que a diferença entre este último e o grupo socioeconómico de médio rendimento não é estatisticamente significativo e que se verifica o contrário na diferença com o grupo socioeconómico de alto rendimento. De acordo com o *odds ratio*, surge um aumento na adesão da escolha da bicicleta elétrica em cerca de 1.7 e 3 vezes, quando se passa do grupo socioeconómico de baixo rendimento para o médio e alto rendimento respetivamente.

Novas TIC vs Escola

Cálculo do *odds ratio* da interacção TIC * Escola:

- Classe de referência: Escola secundária Adolfo Portela (Escola = 1)

Logit 1: Escola = 2 e TIC = 1 e todas as outras covariáveis num dado valor

Logit 0: Escola = 1 e TIC = 1 e todas as outras covariáveis com o mesmo valor do *logit 1*

$$\exp(\text{logit } 1 - \text{logit } 0) = \exp(0.379 + (-0.889)) = 0.6004956 \cong \mathbf{0.60}$$

A interacção entre as novas tecnologias de informação e comunicação e as escolas secundárias Adolfo Portela e Marques de Castilho é estatisticamente significativa, ao nível de significância a 5%. De acordo com o *odds ratio*, os alunos da escola secundária Marques de Castilho desvalorizam a escolha da bicicleta elétrica, na presença das novas TIC, quando comparados com os alunos da escola secundária Adolfo Portela. Ocorre um decréscimo de 40% na adesão da escolha da bicicleta elétrica quando se passa da escola secundária Adolfo Portela para a escola secundária Marques de Castilho, na presença das novas TIC.

Com efeito, é possível obter uma melhor interpretação através das tabelas de contingência relacionando as 2 covariáveis com a variável resposta:

Ausência das novas TIC				
		Escola		
		1	2	
Escolha	0	164	364	528
	1	29	130	159
		193	494	687

Tabela 6.6: Cruzamento da ausência das novas TIC com a Escola

Presença das novas TIC				
		Escola		
		1	2	
Escolha	0	65	170	235
	1	29	80	109
		94	250	344

Tabela 6.7: Cruzamento da presença das novas TIC com a Escola

De acordo com as tabelas de contingência, houve um maior número de observações (respostas válidas) para o caso da ausência das novas TIC do que para o caso da presença da mesma. Para além disso, o facto de existirem mais observações retiradas da escola secundária Marques de Castilho do que na escola secundária Adolfo Portela faz com que a covariável “Escola” seja um factor que interage com as novas TIC. Ao comparar ambas as tabelas, é possível verificar um aumento mais significativo na escolha da bicicleta elétrica por parte dos alunos da escola secundária Adolfo Portela (16%) do que da escola secundária Marques de Castilho (6%), transitando da ausência das novas TIC para a presença da mesma.

Comprimento da Ciclovía vs Escola

Cálculo do *odds ratio* da interacção CompPC * Escola:

- Classe de referência: Escola secundária Adolfo Portela (Escola = 1)

CompPC = 50%

Logit 1: Escola = 2 e CompPC = 50%, todas as outras covariáveis num dado valor

Logit 0: Escola = 1 e CompPC = 50%, todas as outras covariáveis com o mesmo valor do *logit 1*

$$\exp(\text{logit } 1 - \text{logit } 0) = \exp(0.379 + 0.58) = 2.609086 \approx 2.61$$

- Classe de referência: Escola secundária Adolfo Portela (Escola = 1)

CompPC = 100%

Logit 1: Escola = 2 e CompPC = 100%, todas as outras covariáveis num dado valor

Logit 0: Escola = 1 e CompPC = 100%, todas as outras covariáveis com o mesmo valor do *logit 1*

$$\exp(\text{logit } 1 - \text{logit } 0) = \exp(0.379 + 0.772) = 3.161353 \approx 3.16$$

Relativamente à interacção entre os diferentes comprimentos da ciclovía (Ausência de ciclovía, 50% de ciclovía e 100% de ciclovía no trajeto) e as Escolas (Marques de Castilho e Adolfo Portela), a diferença entre a classe de referência (ausência de ciclovía) e a categoria dos 50% de ciclovía não é estatisticamente significativa, contudo já se apresenta estatisticamente significativa para a categoria dos 100% de ciclovía, quando esta interage com a escola secundária Marques de Castilho.

De acordo com o sumário do modelo final obtido, os alunos da escola secundária Marques de Castilho valorizam mais este atributo do que os alunos da escola secundária Adolfo Portela. Para as categorias relativas à existência de 50% e de 100% de ciclovias foi obtido um aumento na adesão da escolha da bicicleta elétrica em cerca de 2.6 e 3.2 vezes por parte dos alunos da escola secundária Marques de Castilho sobre os alunos da escola secundária Adolfo Portela.

Com efeito, é possível obter uma melhor interpretação através das tabelas de contingência relacionando as 2 covariáveis com a variável resposta:

Ausência de Ciclovias				
		Escola		
		1	2	
Escolha	0	76	192	268
	1	18	51	69
		94	243	337

Tabela 6.8: Cruzamento da ausência de ciclovias com a Escola

Existência de 50% de ciclovias				
		Escola		
		1	2	
Escolha	0	77	174	251
	1	22	75	97
		99	249	348

Tabela 6.9: Cruzamento da existência de 50% de ciclovias com a Escola

Existência de 100% de ciclovias				
		Escola		
		1	2	
Escolha	0	76	168	244
	1	18	84	102
		94	252	346

Tabela 6.10: Cruzamento da existência de 100% de ciclovias com a Escola

De acordo com as tabelas de contingência, é possível verificar que cada uma delas se encontra com um número total de observações muito próximo entre si. É possível verificar um aumento de 3% na adesão na escolha da bicicleta elétrica, na escola secundária Adolfo Portela, na passagem da ausência de ciclovias para a existência de 50% de ciclovias no trajeto, mas logo decresce a mesma percentagem quando se passa da existência de 50% para 100% de ciclovias no trajeto. Na escola secundária Marques de Castilho ocorre um aumento na adesão da escolha da bicicleta elétrica na passagem da ausência de ciclovias para a existência de 50% de ciclovias (9%) e na passagem da existência de 50% para 100% de ciclovias (3%).

6.4 Conclusões

Conforme referido no capítulo 1, as questões que a presente investigação pretendeu dar respostas foram:

- Qual o papel futuro na mobilidade elétrica de duas rodas nas deslocações casa/escola?
- Que atributos/características da bicicleta elétrica são mais valorizadas pelas famílias?

Em seguida, resumem-se os principais contributos da presente dissertação na resposta às referidas questões.

Na primeira parte do inquérito foi possível concluir que cerca de 50% dos estudantes utilizam o automóvel na sua deslocação casa/escola, indicando uma forte dependência no uso do referido transporte devido ao seu conforto e ao menor tempo de viagem despendido quando comparado com a bicicleta convencional. De frisar que apenas um inquirido (0.4%) afirmou usar a bicicleta convencional no seu percurso casa/escola. Dos 248 estudantes, cerca de 99% afirmaram saber andar de bicicleta mas apenas 16% afirmam usarem a mesma regularmente e só 6% utilizou a bicicleta elétrica de Águeda. No percurso casa/escola, em bicicleta elétrica, as condições climatéricas (cerca de 53%) e o pavimento irregular (cerca de 40%) foram as barreiras consideradas mais relevantes (apenas 15 inquiridos responderam à questão), enquanto em bicicleta convencional o risco de acidente com outros veículos (33.5%), o relevo (30.5%) e as condições climatéricas (25%) foram as barreiras mais valorizadas pelos inquiridos (foram obtidas 200 respostas à questão). Independentemente do tipo de bicicleta (convencional ou elétrica), os inquiridos consideraram as condições climatéricas a barreira mais relevante e que o relevo deixou de ser considerado um obstáculo quando se mudou da bicicleta convencional para a bicicleta elétrica. As características do transporte elétrico de duas rodas mais valorizadas, pelos inquiridos são a acessibilidade no preço da bicicleta elétrica (cerca de 60%), o esforço reduzido e a segurança rodoviária (26%), revelando uma preocupação sobre os custos e a segurança relativo ao veículo elétrico de duas rodas.

Sobre as escolas, concluiu-se que os estudantes da escola secundária Marques de Castilho usam mais o automóvel nas suas deslocações casa/escola e valorizam menos as barreiras “risco de acidente com outros veículos” e “ausência de conforto” do que os estudantes da escola secundária Adolfo Portela.

Quanto ao género, o grupo do género masculino usa mais o automóvel do que o grupo do género feminino. Por outro lado, a barreira “tráfego motorizado na infraestrutura (VLP) é mais valorizada pelo grupo do género feminino. Quanto às características da bicicleta elétrica, o grupo do género masculino valoriza mais o conforto da suspensão enquanto o grupo do género feminino valoriza mais o esforço físico reduzido e a segurança pessoal.

Na 2ª parte do inquérito foram abordadas questões sobre preferências, por parte dos inquiridos, acerca do regime de utilização da bicicleta elétrica adotado pelas famílias, o modelo favorito a ser usado e o sistema de equipamento tecnológico a adicionar ao seu veículo elétrico de duas rodas. Foi concluído que a maioria das famílias preferiu optar pela utilização da bicicleta convencional (cerca de 50%) tendo como alternativa a compra (cerca de 14%) ou aluguer (cerca de 17%) da bicicleta elétrica. Entre os modelos apresentados no inquérito, o modelo “Montanha/Desporto” foi o mais votado pelas famílias (cerca de 52%), seguido do modelo “utilitário” (cerca de 26%) e do modelo “design personalizado” (cerca de 21%). Relativamente à inserção de sistemas tecnológicos, a maioria dos inquiridos (50%) preferiu adicionar o sistema de TIC (GPS), seguido dos 4 equipamentos em simultâneo (43%) e do sistema de contabilização de calorias (25%).

De acordo com as escolas, os estudantes da escola secundária Adolfo Portela preferem comprar a bicicleta elétrica enquanto os alunos da escola secundária Marques de Castilho preferem alugar a mesma.

Concluiu-se, quanto ao género, que o grupo do género feminino prefere o modelo utilitário, enquanto o grupo do género masculino prefere o modelo “montanha/desporto” ou o modelo

“design personalizado”. Relativamente aos sistemas adicionais, o grupo do género masculino valoriza mais o sistema cooperativo de bicicletas do que o grupo do género feminino.

De acordo com o grupo socioeconómico, o grupo de baixo rendimento prefere usar a bicicleta convencional, enquanto os grupos de médio e alto rendimento preferem não optar por nenhum regime. Quanto aos modelos de bicicleta elétrica, o grupo socioeconómico de baixo rendimento prefere o modelo “Montanha/Desporto”, enquanto os grupos socioeconómicos de médio e alto rendimento preferem o modelo utilitário.

A estimação do modelo de regressão logística permitiu identificar quais os fatores mais significativos e influentes no sentido da adesão dos alunos à bicicleta elétrica. No modelo foram identificadas 9 covariáveis associadas à variável resposta (escolha do automóvel ou da bicicleta elétrica), nomeadamente a diferença de custo entre a bicicleta elétrica e o automóvel, o comprimento da ciclovia, a inserção das novas TIC, a diferença do tempo de viagem entre a bicicleta elétrica e o automóvel, a escola, a experiência anterior em bicicleta, o grupo socioeconómico, a interação entre as novas TIC e a escola e a interação entre o comprimento da ciclovia e a escola. As covariáveis enunciadas estando associadas à variável resposta, podem promover uma das duas categorias representativas da mesma.

Foi diagnosticado no caso estudo que as covariáveis que estão associadas à adesão da escolha do automóvel são: a diferença de custo menor (Custo mais caro para a bicicleta elétrica), a diferença de tempo de viagem maior (tempo de viagem mais longo para a bicicleta elétrica), o grupo socioeconómico de baixo rendimento e os alunos da escola secundária Marques de Castilho na presença das novas TIC.

As covariáveis que contribuíram na adesão da escolha da bicicleta são: a existência de 100% de comprimento de ciclovia, a presença das novas TIC, a experiência anterior em bicicleta, os alunos da escola secundária Marques de Castilho e os alunos da escola secundária Marques de Castilho com a existência de 100% ou mesmo de 50% de comprimento de ciclovia.

De seguida descrevem-se os efeitos de cada covariável na escolha do veículo e, ainda as possíveis justificações para o seu efeito esperado ou característico sobre a variável resposta, de acordo com a literatura existente, isto é, sobre se o efeito resultante é no sentido da escolha da opção do automóvel ou da bicicleta elétrica.

Custo

Numa situação real de procura e oferta de um determinado veículo, é de esperar que quanto maior for o seu custo, menor é a adesão à sua escolha. Assim, verifica-se uma diminuição na adesão à escolha da bicicleta elétrica à medida que o custo desta vai aumentando. Esta situação pode estar ligada ao comportamento receoso por parte do consumidor quanto ao preço elevado de aquisição. No caso estudo onde se tem de decidir entre um veículo elétrico de duas rodas e um automóvel movido a combustível surge também um problema relativo ao comportamento receoso por parte do consumidor quanto ao preço elevado de aquisição de uma bicicleta elétrica.

Comprimento da Ciclovia

Neste estudo foi possível observar que a existência de 100% de uma ciclovia no trajeto provoca um aumento significativo na adesão à escolha da bicicleta elétrica. De acordo com alguns estudos efectuados noutros países, uma das possíveis explicações para tal aumento pode estar relacionada com a perceção do risco de acidente no caso de partilha da infraestrutura rodoviária, uma vez que as bicicletas convencionais/elétricas não conseguem acompanhar a velocidade usual dos automóveis. Para além disso, as consequências mais graves, em caso de acidente, recaem com maior peso sobre os condutores de bicicleta elétrica do que de automóvel, devido a questões como a instabilidade da condução da bicicleta elétrica (como de qualquer veículo de duas rodas) e da falta de recursos para a segurança do próprio condutor desse veículo [*Mobility and Transport, European Road Safety Laboratory, 2012*].

Por outro lado, surgiram estudos que afirmaram que a presença de bicicletas elétricas nas ciclovias pode provocar acidentes graves sobre os peões ou os condutores de bicicletas convencionais, uma vez que os veículos elétricos atingem velocidades na ordem dos 60km/h (muito superior à velocidade atingida pela bicicleta convencional). Para este caso tem-se então uma controvérsia que necessita de ser estudada com mais pormenor [Wei Du et al, 2013].

Em países como a China, onde a maior parte da população circula de bicicleta convencional, ocorre o problema particular de lotação de tráfego nas grandes cidades, aumentando vertiginosamente o número de acidentes envolvendo bicicletas elétricas e automóveis ou bicicletas convencionais.

Na Holanda ou na Dinamarca foi obtido um enorme sucesso na implementação da bicicleta elétrica ou mesmo de bicicleta convencional pois as ruas das cidades de Bruges, Amsterdão e Copenhaga são, na sua maioria, planas e as suas infraestruturas foram levadas a cabo de uma forma eficaz sobre o alisamento do caminho das ciclovias. Para além das infraestruturas foram também criados vários impostos e restrições no uso do automóvel, sendo parte desses impostos utilizados precisamente nas infraestruturas das bicicletas e não dos automóveis. Na Dinamarca são removidas cerca de 3% de vagas para carros todos os anos, transformando-os num espaço público ou de lazer, fazendo com que circular de carro seja impraticável e apenas útil em viagens de longa distância.

Novas TIC

Com a evolução significativa da tecnologia é de esperar que a inserção das novas TIC produza um efeito positivo na adesão da bicicleta elétrica, dado que também se havia concluído na análise exploratória a existência de uma elevada percentagem revelando uma visível necessidade de incluir as novas TIC nas bicicletas elétricas (53%).

Para além disso, foram analisadas as preferências sobre 4 equipamentos adicionais para a bicicleta elétrica: Sistema de TIC, Sistema de contabilização de calorias consumidas na deslocação, Sistema de tecnologia Bluetooth e o Sistema cooperativo de bicicletas. Foi dada uma maior preferência sobre o sistema de TIC (possivelmente por ser o equipamento mais usual e, por esse motivo, mais conhecido entre os inquiridos) e sobre o sistema de contabilização de calorias (mais direcionado para os inquiridos que pretendam usar este meio de transporte como meio de praticar desporto).

Foi também concluído que o grupo do género masculino revela uma maior preferência sobre a inserção de equipamentos adicionais do que o grupo do género feminino, nomeadamente do sistema cooperativo de bicicletas ou os 4 equipamentos em simultâneo.

Tempo de viagem

De acordo com o caso estudo, verificou-se uma diminuição na adesão da escolha da bicicleta elétrica com o aumento do tempo de viagem desta sobre o tempo de viagem, para o mesmo trajeto, efectuado pelo automóvel. De um modo geral, a bicicleta, mesmo sendo elétrica, demora mais tempo a chegar ao destino do que o automóvel para um mesmo trajeto, sendo de esperar um efeito de despromoção na escolha da bicicleta elétrica.

Escola

Quanto às escolas secundárias foi observado um aumento na promoção da escolha da bicicleta elétrica na transição da escola secundária Adolfo Portela para a escola secundária Marques de Castilho. Em particular, esta variável, apesar de não ser estatisticamente significativa, permite concluir que os alunos da escola secundária Marques de Castilho promovem a escolha da bicicleta elétrica, enquanto os alunos da escola secundária Adolfo Portela desvalorizam a promoção da mesma.

Apesar de serem os alunos da escola secundária Marques de Castilho aqueles que possuem uma maior dependência no uso do automóvel, de acordo com a análise exploratória do caso estudo, e, em simultâneo, também apresentarem uma maior tendência na promoção da

escolha da bicicleta elétrica, pode estar relacionado com o elevado custo de aquisição do meio de transporte elétrico de duas rodas. Desta forma, este meio de transporte é apenas viável para os indivíduos pertencentes ao grupo socioeconómico de alto rendimento (utilizadores do automóvel).

Experiência anterior em bicicleta

Relativamente à experiência anterior em bicicleta, foi possível concluir que quanto maior for a experiência do aluno maior é a contribuição no sentido de promover a escolha da bicicleta elétrica. É de esperar que quanto maior for a experiência anterior em bicicleta, maior é a segurança e a confiança em querer utilizar/experimentar a bicicleta elétrica.

Grupo Socioeconómico

Através do caso estudo foi possível concluir que quanto menor for o rendimento do grupo socioeconómico, menor é a adesão na escolha da bicicleta elétrica. Esta situação vem servir de suporte à informação fornecida pela análise exploratória que revela que o grupo socioeconómico de baixo rendimento prefere usar a bicicleta convencional, enquanto os restantes grupos sócio económicos preferem não optar por nenhum regime. Por outro lado, o grupo socioeconómico de baixo rendimento, no que diz respeito aos modelos favoritos da bicicleta elétrica, elegem o modelo montanha/desporto como o seu modelo favorito enquanto os restantes grupos sócio económicos valorizam o modelo utilitário. De acordo com os favoritismos dos modelos da bicicleta elétrica em função dos grupos sócio económicos, o grupo socioeconómico de baixo rendimento visa utilizar a bicicleta como um meio de transporte de lazer e de esporádica utilização, enquanto os grupos socioeconómico de médio e alto rendimento procuram na bicicleta um meio de transporte prático e de uso regular.

Em países como na China, mais especificamente na cidade de Shangai, o custo de aquisição de uma bicicleta elétrica produz o efeito contrário ao que é verificado no caso estudo presente, uma vez que este meio de transporte possui um preço de aquisição muito mais acessível. Por este motivo, os utilizadores de bicicletas elétricas, na sua maioria, pertencem aos grupos sócio económicos de baixo e médio rendimento.

Novas TIC vs Escola

Na transição da ausência das novas TIC para a presença das mesmas, ocorre um aumento mais significativo na escolha da bicicleta elétrica por parte dos alunos da escola secundária Adolfo Portela do que por parte dos alunos da escola secundária Marques de Castilho.

Com efeito, já havia sido concluído que os alunos da escola secundária Marques de Castilho possuem uma maior dependência no transporte automóvel no seu trajeto casa/escola do que os alunos da escola secundária Adolfo Portela. Para além disso a preferência da presença das novas TIC por parte dos alunos da escola secundária Adolfo Portela pode ser explicada por esta apresentar projetos inovadores ligados à Ciência e cursos avançados ligados a áreas tanto científicas como às novas tecnologias de comunicação. Desta forma, a escola secundária Adolfo Portela é representativa de alunos com interesse, na sua maioria, nas áreas ligadas às novas tecnologias.

Comprimento da Ciclovía vs Escola

Na passagem da ausência de ciclovía para a existência de 100% de ciclovía, houve um aumento de 12% na adesão da escolha da bicicleta elétrica por parte dos alunos da escola secundária Marques de Castilho enquanto na escola secundária Adolfo Portela não houve qualquer alteração.

Com efeito, de acordo com a análise exploratória, foi concluído que os alunos da escola secundária Marques de Castilho valorizam mais as barreiras “Risco de acidentes com peões”, “Relevo” e “Pavimento irregular” na sua deslocação casa/escola, revelando uma forte preocupação sobre as condições de uma infraestrutura para a circulação das bicicletas.

A análise estatística efetuada no presente projeto teve por base a utilização de dois *softwares*, *SPSS* e *R studio*, na análise exploratória e na modelação, respectivamente. Através do *SPSS* foram aplicados sobre os dados tabelas cruzadas com vista a obter possíveis associações entre as covariáveis, por outro lado, de acordo com o *R studio*, foi possível estimar um modelo com todas as covariáveis significativas para o mesmo, permitindo avaliar, deste modo, aquelas que produzem uma tendência positiva na adesão à bicicleta elétrica.

Dado que em 1031 observações válidas obtidas na análise das 9 opções de mobilidade, 268 respostas a favor da bicicleta elétrica e 763 a favor do automóvel, significa que ainda existe uma grande dependência no uso do automóvel e consequentemente existe uma grande contenção na mudança para o veículo elétrico de duas rodas. O presente estudo mostra que as novas TIC podem ajudar a aumentar a atratividade da bicicleta elétrica para o segmento de mercado analisado, viabilizando assim o seu mercado potencial no futuro. De acordo com a análise efetuada ao longo do projeto, são recomendadas algumas linhas de investigação futura na área da estatística nomeadamente, a aplicação de métodos adequados para o tratamento de dados omissos, teste/estimação de outro tipo de modelos (e.g. *random parameters logit*) que tenha em consideração as diferentes preferências de vários grupos de potenciais utilizadores da bicicleta elétrica e uma recolha de dados sobre inquiridos que realmente utilizassem a bicicleta convencional/elétrica com o fim de se perceber qual a sua interação com o restante tráfego.

Bibliografia

1. Gartner, John and Citron, Ryan; *Research Report, Executive Summary: Electric Bicycles - Throttle-Control and Pedal-Assist E-Bicycles, Batteries, and Motors: Global Market Opportunities, Barriers, Technology Issues and Demand Forecasts*, 2014, Navigant Research;
2. Marques da Silva, José Carlos (2013), Sistema de mobilidade elétrica de duas rodas de elevada eficiência e performance, Dissertação de Mestrado, Engenharia Electrotécnica - Energia e Automação Industrial, Instituto Politécnico de Viseu;
3. Associação Portuguesa do Veículo Elétrico, MOBIL3e, Mobilidade Elétrica, Eficiência Energética, 2012 APVE;
4. Câmara Municipal de Lisboa, MOBI.E-Rede de Mobilidade Elétrica, 2014;
5. Wikipédia, a enciclopédia livre: Mobilidade Sustentável;
6. Wikipédia, a enciclopédia livre: Rede Elétrica Inteligente;
7. Mobility and Transport, European Road Safety Laboratory, Candappa N., et al. (2012) Basic Fact Sheet "Cyclists", Deliverable D3.9 of the EC FP7 project DaCoTA;
8. MOBI.E - Mobilidade Elétrica, PT, INTELI, Inteligência em Inovação 2010;
9. Correio Braziliense: Ciência e Saúde, Outubro de 2013, Obesidade atinge quase 35% da população adulta dos Estados Unidos, Brasília;
10. Gunther, Marc, *Yale Environment 360: "Will Electric Bicycles Get Americans to Start Pedaling?"*, Abril de 2013, *Business & Innovation Climate Policy & Politics Policy & Politics Urbanization Asia and North America*;
11. Portal do Ambiente e do Cidadão: "Sobre a Mobilidade em Portugal e na Europa", 2005, Departamento do Ambiente e Qualidade de Vida - Câmara Municipal da Maia;
12. Mariño Pequini, Suzi (2000), A Evolução Tecnológica da Bicicleta e suas Implicações Ergonómicas para a Máquina Humana: "Problemas da Coluna Vertebral x Bicicletas dos Tipos *Speed* e *Mountain Bike*" - Capítulo 2: Evolução Histórica da Bicicleta, 2000;
13. Kang, An et al (2013), *Travel Characteristics of E-bike Users: Survey and Analysis in Shanghai*, *Procedia Social Behavioral Sciences* 96, p 1828-1838;
14. Bai, Lu et al (2013), *Comparative Analysis of the safety effects of electric bikes at signalized intersections*, *Transportation Research Part D* 20, p 48-54;
15. Popovich, Natalie et al (2014), *Experiences of electric bicycle users in the Sacramento, California area*, *Travel Behaviour and Society* 1, p 37-44;

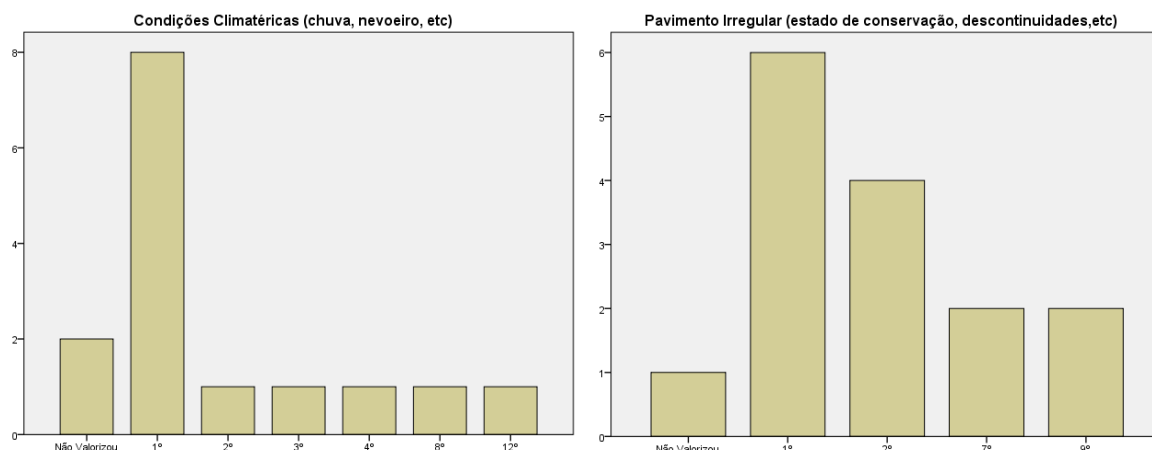
16. Wei, Longyu et al (2013), *Comparison Study on Travel Characteristics between Two Kinds of Electric Bike*, *Social Behavioral Sciences* 96, p 1603-1610;
17. Du, Wei et al (2013), *Understanding on-road practices of electric bike riders: An observational study in a developed city in China*, *Accident Analysis and Prevention* 59, p 319-326;
18. Li, Zhibin et al (2012), *Physical environments influencing bicylists perception of comfort on separated and on-street bicycle facilities*, *Transportation Research Part D* 17, p 256-261;
19. Cycletourer website: *Cycling in Germany* (2014), *Bycicle Traveler*;
20. Wikipédia, a enciclopédia livre: *German Cycling Network*;
21. Wikipédia, a enciclopédia livre: *Electric bicycle*;
22. Wikipédia, a enciclopédia livre: *Cycling in Denmark*;
23. Wikipédia, a enciclopédia livre: *Cycling in the Netherlands*;
24. Wikipédia, a enciclopédia livre: *Electric bicycle laws*;
25. McCullagh, P. e Nelder, John A. (1989) *Monographs on Statistics & Applied Probability -Generalized linear models*. Chapman and Hall_CRC;
26. Dobson, Annette J. (2001). *An Introduction to Generalized Linear Models*, Second Edition-Chapman & Hall;
27. Gilberto A. Paula, (2004). *Modelos de Regressão com apoio computacional*, Instituto de Matemática e Estatística da Universidade de São Paulo;
28. McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd Edition. London: Chapman and Hall;
29. Hardin, J. and Hilbe, J. (2012). *Generalized Linear Models and Extensions*, 3rd Edition. College Station, Texas: Stata Press;
30. Felgueiras, Miguel M. (2003). *Análise de Dados Binários*. Tese de Mestrado. Faculdade de Ciências da Universidade de Lisboa;
31. Turkman, M. A. e Silva, G. L. (2000). *Modelos Lineares Generalizados*. SPE;
32. Hosmer, D. W. e Lemeshow, S. (2000). *Applied Logistic Regression*. 2nd Edition, Wiley Series in Probability and Statistics.
33. Collett, D. (2003). *Modelling Binary Data*, 2nd Edition, Texts in Statistical Science, Chapman & Hall;
34. Hosmer, D. W., Hosmer, T., Le Cessie, S., e Lemeshow, S. (1997). *A Comparison of goodness of fit tests for the logistic regression model*. *Statistics in Medicine*, Vol. 16, 965-980;
35. Esarey, Justin e Pierce, Andrew (2012). *Assessing fit quality and testing for misspecification in Binary Dependent Variable Models*;
36. McFadden, Daniel (1974). *Conditional logit analysis of qualitative choice behavior*. *FRONTIERS IN ECONOMETRICS*, Cap. 4, 105-142, Academic Press: New York;
37. Berridge, Damon M. e Crouchley, Robert (2011). *Multivariate Generalized Linear Mixed Models using R*. Chapman and Hall CRC;
38. Barbieri, N. B. (2012). *Estimação Robusta para o Modelo de Regressão Logística*. Tese de Bacharelato. Universidade Federal do Rio Grande do Sul;
39. Box, G. E. P. e Cox, D. R. (1964). *An analysis of Transformations*. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 26, No. 2. p. 211-252;
40. Kern, Holdger e Grimmer, Justin (2003). *How to obtain confidence intervals without simulation: The Delta Method*;
41. De Oliveira, Samuel (2013). *Inferência e Análise de Resíduos e de Diagnóstico em Modelos Lineares Generalizados*. Tese de Mestrado. Universidade Federal de Juiz de Fora;
42. Hu, Bo., Shao, Jun. e Palta, Mari. (2006) *Pseudo-R² in Logistic Regression Model*. *Statistica Sinica* 16, 847-860;
43. F. de Oliveira, Nelson; S. Santana, Vilma e Alberto Lopes, Antonio. (1997) *Razões de Proporções e uso do Método Delta para intervalos de Confiança em Regressão Logística*. *Revista de Saúde Pública*, Vol. 31, nº1, 90-99;
44. Cochran, William G. (1977), *Sampling Techniques*, 3rd edition, John Wiley and Sons;

45. Barnett, V. (1974)- *Elements of Sampling Theory*, English Univ. Press;
46. B. S. Everitt (1977), *The Analysis of Contingency Tables*, 1st edition, Springer US;
47. Longnecker, M.T & Ott, R.L : *A First Course in Statistical Methods*, page 615. Thomson Brooks/Cole, 2004;
48. Studenmund, A.H: *Using Econometrics: A practical guide*, 5th Edition, page 258-259. Pearson International Edition, 2006;
49. Hair JF, Anderson R, Tatham RL, Black WC: *Multivariate Data Analysis*. Prentice Hall: Upper Saddle River, N.J. 2006;
50. Marquardt, D.W. 1970 *Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation*, *Technometrics* 12(3), 591, 605-07;
51. Allison, P.D. *Multiple Regression: a primer*, page 142. Pine Forge Press: Thousand Oaks, C.A. 1999;
52. Kutner MH, Nachtsheim CJ, Neter J (2004), *Applied Linear Regression Models*, 4th edition, McGraw-Hill Irwin;
53. Hanley, J.A., and McNeil, B. J. (1983). *Radiology* 148:839-843;
54. Bindslev Niels (2008), *Drug-Acceptor Interactions - Modeling theoretical tools to test and evaluate experimental equilibrium effects*;
55. Fawcett, Tom (2006). *An Introduction to ROC Analysis*, *Pattern Recognition Letters* 27 (8): 861-874;
56. Rajul Parikh, MS, Annie Mathai, MS, Shefali Parikh, MD, G Chandra Sekhar, MD, and Ravi Thomas, MD, *Understanding and using sensitivity, specificity and predictive values*, *Indian Journal of Ophthalmology*, 2008 Jan-Feb, 45-50;
57. Mason, Simon J.; Graham, Nicholas E. (2002). *Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation*, *Quarterly Journal of the Royal Meteorological Society* (128): 2145-2166;
58. Robert Taylor, John (1999). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books. pp. 128-129;
59. Arsenio, E. (2013) - Estudo piloto be4schools, LNEC, Setembro de 2013;
60. *Electric Bicycle Guide* (2010-2013), "What's the difference between pedelecs and e-bikes?";
61. *MotorDream* (2014), Roda especial transforma bicicletas comuns em elétricas;
62. "Resolução do Conselho de Ministros n.º81/2009," *Diário da República*, 1.ª série – N.º 173, 7 setembro 2009,
63. *Science of Cycling: Human Power* [online];
64. "Decreto-lei n.º 170/2012", *Diário da República*, 1.ª série - N.º 148, 1 de Agosto de 2012;
65. REA - Relatório do Estado do Ambiente 2014 Portugal, APA - Agência Portuguesa do Ambiente 2014, Ministério do Ambiente, ordenamento do território e energia;
66. Estatísticas dos Transportes 2011, INE - Instituto Nacional de Estatística, Edição 2012;
67. Eurostat: Database - Transport, European Statistics.
68. Andreína Vieira de Abreu, Daisy (2013), *Síndrome Coronário Agudo: Análise do impacto das variáveis sociodemográficas, ambientais e clínicas na demora média entre o início da sintomatologia e o restabelecimento do fluxo*, Dissertação de Mestrado em Bioestatística, Faculdade de Ciências da Universidade de Lisboa.

Apêndice 1

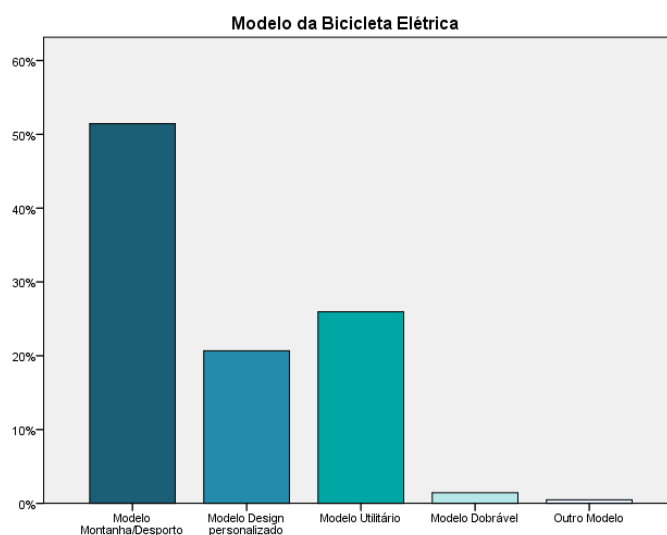
1. Percepção de barreiras na sua deslocação casa/escola, em bicicleta elétrica

De acordo com a sua percepção, que barreiras (obstáculos) existem?



De acordo com as figuras acima, as condições climáticas e o pavimento irregular foram as barreiras consideradas de maior importância por parte dos inquiridos.

2. Modelo da bicicleta elétrica

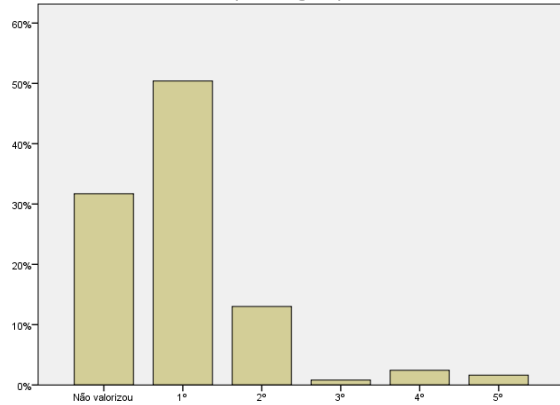


De acordo com a figura acima, o modelo montanha/desporto é o preferido pela maioria dos inquiridos (mais de 50%), seguido do modelo utilitário e do modelo de design personalizado.

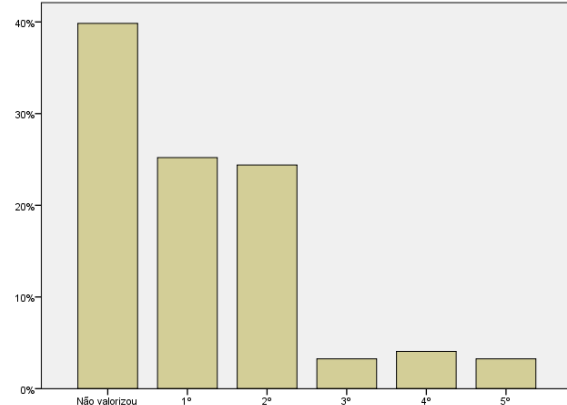
3. Inserção de um sistema de equipamento de novas tecnologias TIC

3.1 Equipamento adicional

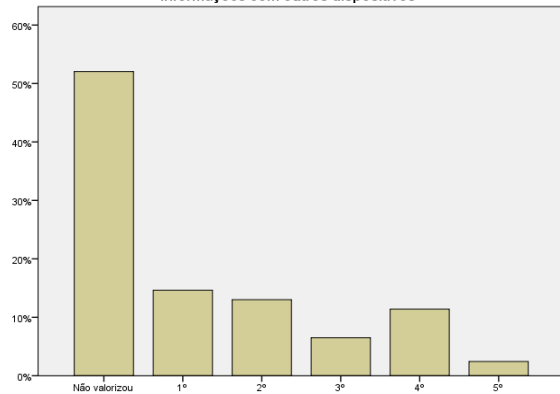
Equipamento Adicional: Sistema de Tecnologia de informação e comunicação (GPS integrado)



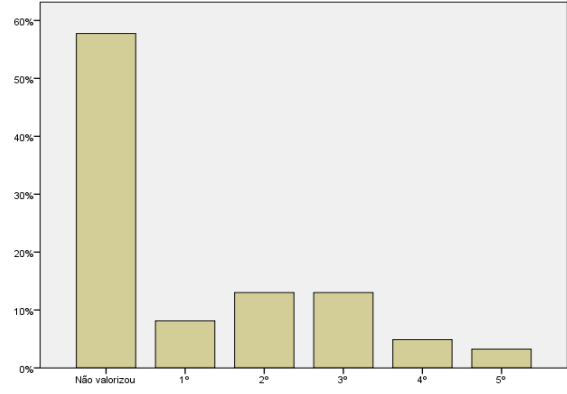
Equipamento Adicional: Sistema de Contabilização de Calorias consumidas na deslocação, em função do esforço físico



Equipamento Adicional: Sistema de Tecnologia Bluetooth que permite trocar informações com outros dispositivos



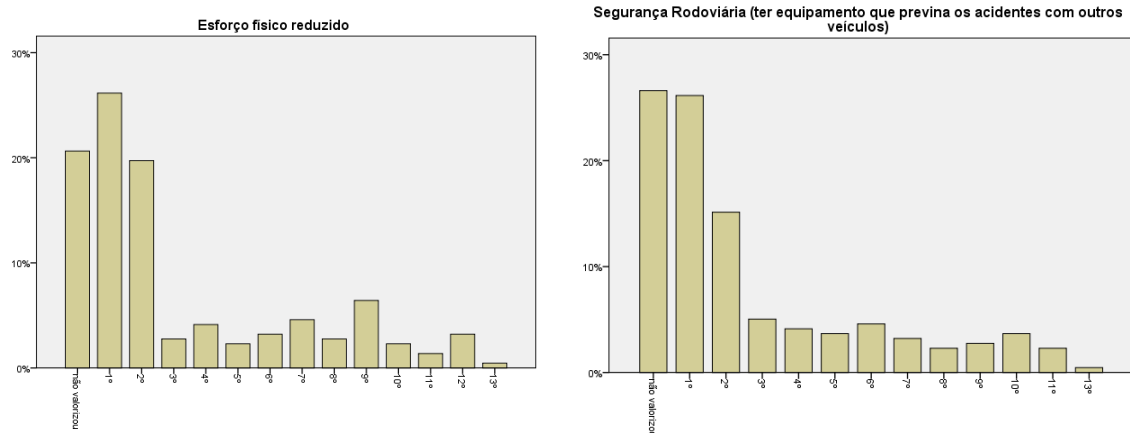
Equipamento Adicional: Sistema Cooperativo de Bicicletas, visando constituir uma rede social application "e-bikers"



Relativamente à inserção de um sistema de equipamento de novas tecnologias, os inquiridos preferem o sistema de tecnologia de informação e comunicação (GPS, cerca de 50%), seguido dos 4 sistemas em simultâneo e do sistema de contabilização de calorias (cerca de 40%).

4. Características da bicicleta elétrica

Quais dos seguintes atributos considera mais importantes na bicicleta elétrica?



Tendo em conta os atributos na bicicleta elétrica que foram apresentados no inquérito, os estudantes consideraram o esforço físico reduzido e a segurança rodoviária os atributos de maior importância no meio de transporte elétrico de duas rodas.

5. Identificação da Escola

5.1 Uso do Automóvel na deslocação casa/escola

Automóvel * Escola Tabulação cruzada

			Escola		Total
			ESAP - Escola Secundária Adolfo Portela	ESMC - Escola Secundária Marques de Castilho	
Automóvel	Não usa	Contagem	24	91	115
		Resíduos ajustados	-1,9	1,9	
	Usa	Contagem	39	85	124
		Resíduos ajustados	1,9	-1,9	
Total		Contagem	63	176	239

De acordo com a tabela de contingência, os resíduos ajustados apresentam valores algo significativos e é possível observar que a maioria dos alunos da escola secundária Marques de Castilho não usa o automóvel como meio de transporte casa/escola enquanto grande parte dos alunos da escola secundária Adolfo Portela usam este meio de transporte no seu trajeto casa/escola.

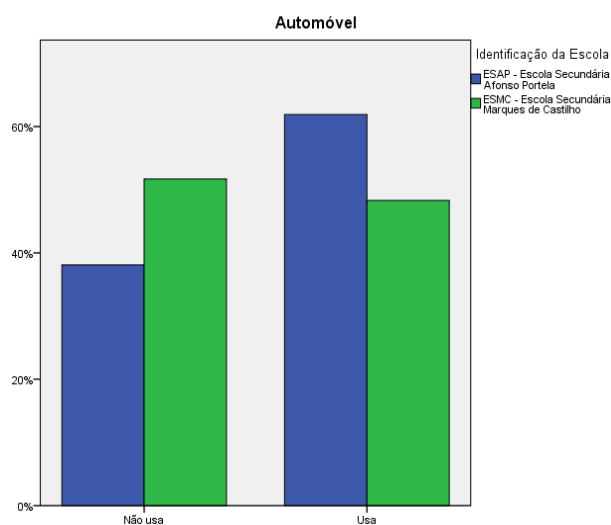
Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)	Sig exata (2 lados)	Sig exata (1 lado)
Qui-quadrado de Pearson	3,442 ^a	1	,064	,078	,043
Correção de continuidade ^b	2,918	1	,088		
Razão de verossimilhança	3,472	1	,062		
Teste Exato de Fisher					
Associação Linear por Linear	3,428	1	,064		
N de Casos Válidos	239				

a. 0 células (,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 30,31.

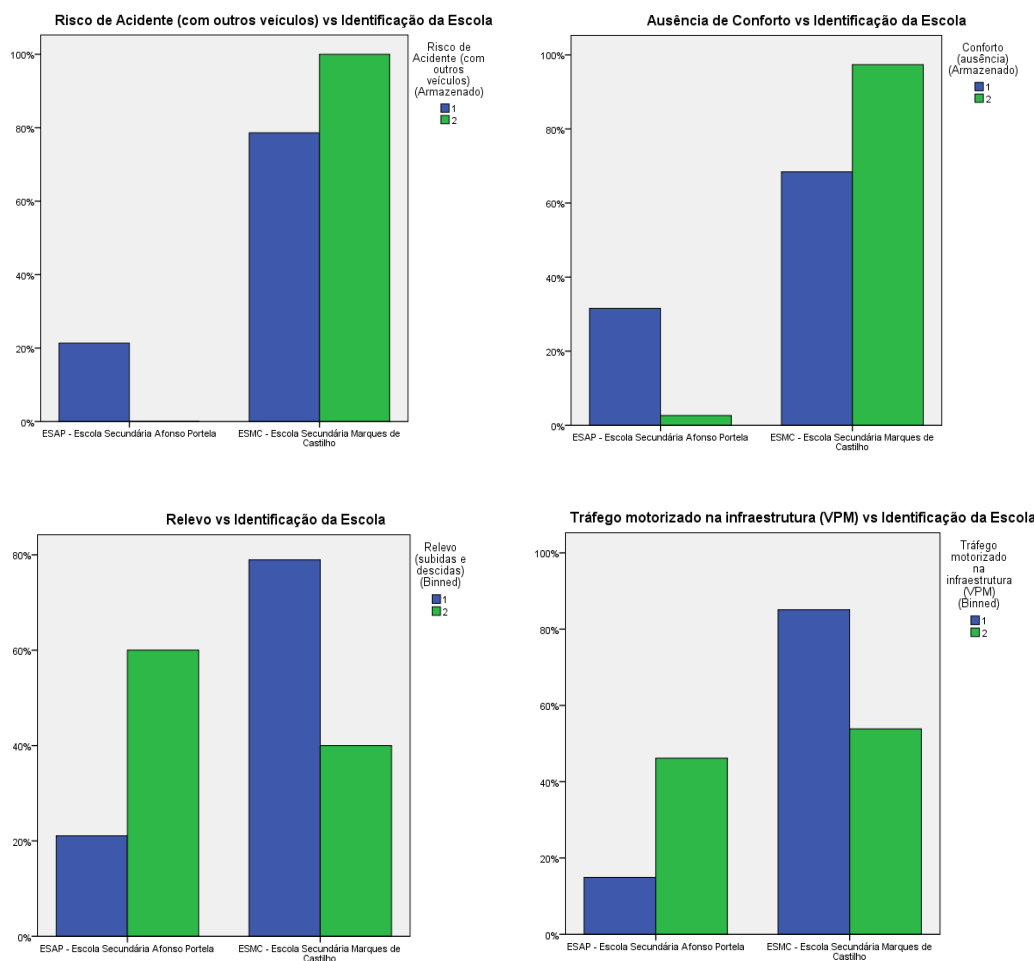
b. Computado apenas para uma tabela 2x2

Uma vez que não existem frequências inferiores a 5 na tabela de contingência, o teste qui quadrado é o mais indicado na análise de uma possível associação entre o uso do automóvel e as escolas secundárias. Tendo em conta o valor p associado ao teste qui quadrado (0.064), rejeitamos a hipótese da existência de independência entre o uso do automóvel e as escolas ao nível de significância de 0.10, isto é, é possível afirmar a existência de uma associação entre o uso do automóvel e as escolas. Assim e de acordo com teste, a associação verificada na tabela de contingência é significativa.



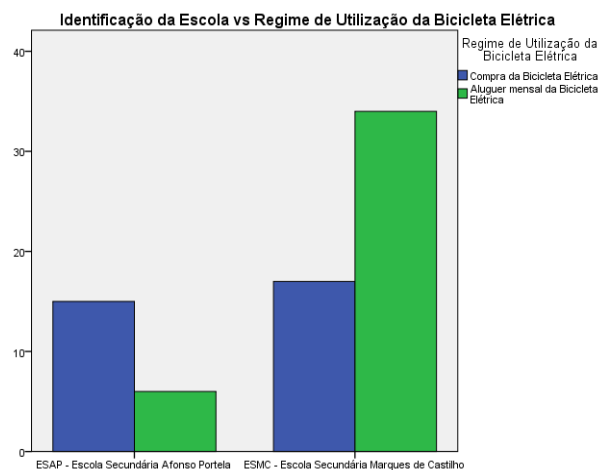
De acordo com a figura anterior, os estudantes da escola secundária Adolfo Portela usam mais o automóvel do que os estudantes da escola secundária Marques de Castilho.

5.2. Percepção de barreiras na sua deslocação casa/escola, em bicicleta convencional



De acordo com as figuras acima, os estudantes da escola secundária Adolfo Portela consideram o risco de acidente com outros veículos e a ausência de conforto as barreiras mais importantes na sua deslocação casa/escola, em bicicleta convencional, do que os estudantes da escola secundária Marques de Castilho. Por outro lado, os estudantes da escola secundária Marques de Castilho valorizam mais as barreiras relacionadas com o relevo e o tráfego motorizado na infraestrutura (veículos pesados de mercadorias) do que os estudantes da escola secundária Adolfo Portela.

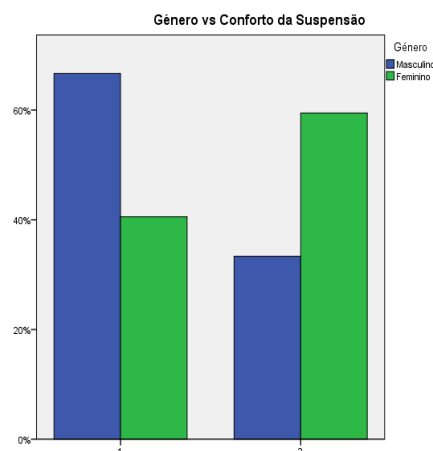
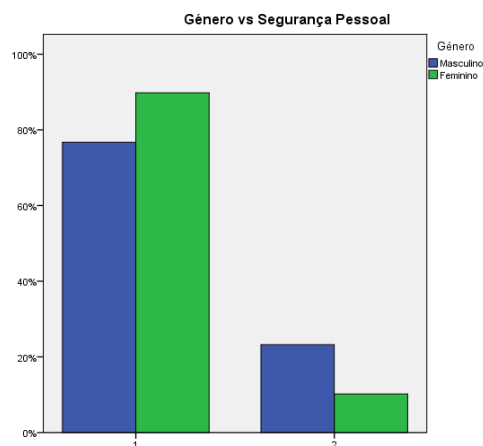
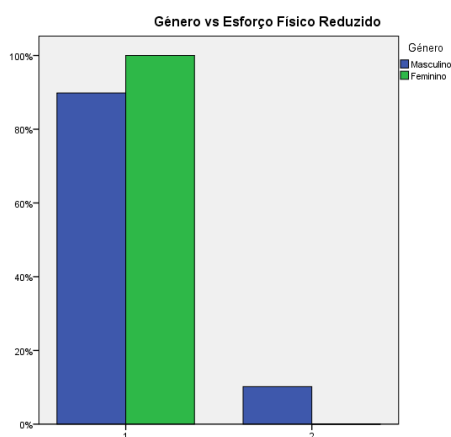
5.3. Regime de utilização da bicicleta elétrica



Relativamente ao regime de utilização da bicicleta elétrica, os alunos da escola secundária Marques de Castilho preferem pagar um aluguer mensal pela bicicleta elétrica enquanto os alunos da escola secundária Adolfo Portela preferem efectuar a compra da mesma.

6. Género

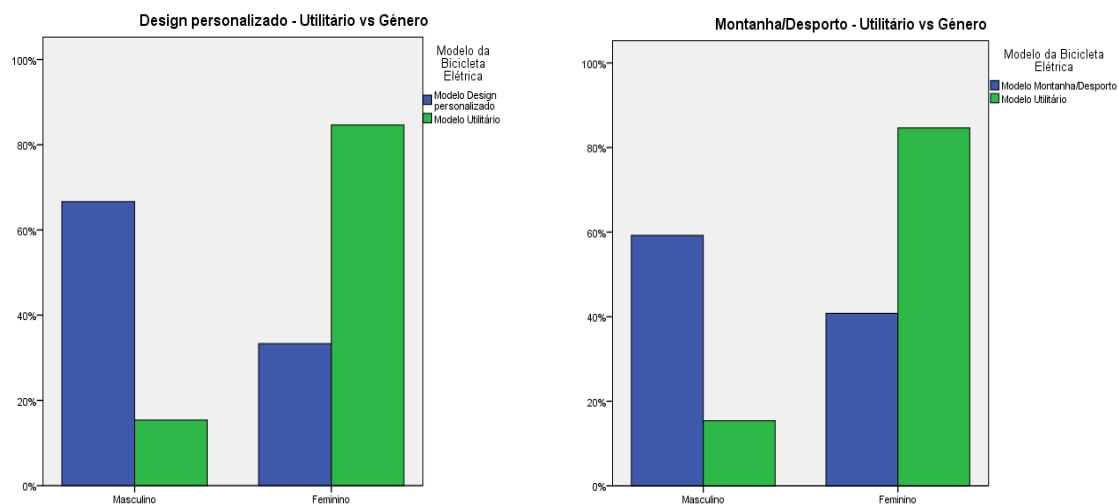
6.1. Características da bicicleta elétrica consideradas importantes



Segundo as características da bicicleta elétrica consideradas importantes (1 - Mais valorizado e 2 - Menos Valorizado), o grupo do género feminino valoriza mais o esforço físico reduzido e

a segurança pessoal, enquanto o grupo do género masculino valoriza mais o conforto da suspensão.

6.2. Modelos da bicicleta eléctrica



De acordo com os modelos da bicicleta eléctrica, o grupo do género feminino prefere o modelo utilitário, enquanto o grupo do género masculino prefere o modelo montanha/desporto ou o modelo design personalizado.

6.3. Equipamento adicional: Sistema Cooperativo de Bicicletas

Género * Equipamento Adicional: Sistema Cooperativo de Bicicletas, visando constituir uma rede social application "e-bikers" Tabulação cruzada

			Equipamento Adicional: Sistema Cooperativo de Bicicletas, visando constituir uma rede social application "e-bikers" (Armazenado)		Total
			1 – Mais valorizado	2 - Menos valorizado	
Género	Masculino	Contagem	8	11	19
		Contagem Esperada	4,6	14,4	19,0
		Resíduos ajustados	2,5	-2,5	
	Feminino	Contagem	2	20	22
		Contagem Esperada	5,4	16,6	22,0
		Resíduos ajustados	-2,5	2,5	
Total		Contagem	10	31	41
		Contagem Esperada	10,0	31,0	41,0

De acordo com a tabela de contingência, os resíduos ajustados apresentam valores significativos e o grupo de alunos do género masculino valoriza mais o sistema cooperativo de bicicletas do que o grupo de alunos do género feminino. No entanto, é necessário garantir se

esta associação é estatisticamente significativa e para esse efeito efectuou-se um teste de independência.

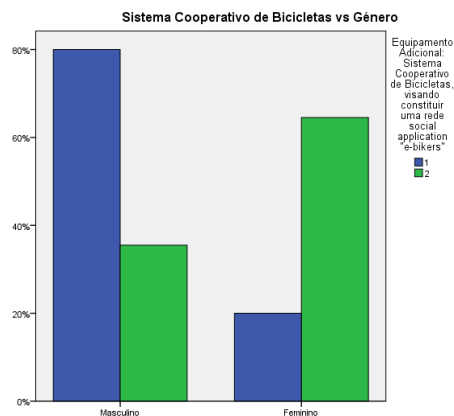
Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)	Sig exata (2 lados)	Sig exata (1 lado)
Qui-quadrado de Pearson	6,026 ^a	1	,014	,026	,017
Correção de continuidade ^b	4,368	1	,037		
Razão de verossimilhança	6,286	1	,012		
Teste Exato de Fisher					
Associação Linear por Linear	5,879	1	,015		
N de Casos Válidos	41				

a. 1 células (25,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 4,63.

b. Computado apenas para uma tabela 2x2

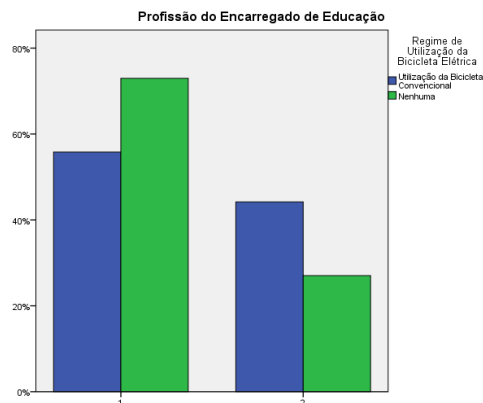
Uma vez que existe mais de 20% das células com uma frequência inferior a 5, o teste a ser utilizado será o teste exacto de Fisher. De acordo com o teste, para um nível de significância de 5%, rejeita-se a hipótese de independência entre o género e o equipamento adicional relativo ao sistema cooperativo de bicicletas. Ao nível de significância a 1% já não é possível afirmar a existência de uma associação entre estes grupos. Assim, a associação existente entre o género dos alunos e o sistema cooperativo de bicicletas é estatisticamente significativo.



O grupo do género masculino valoriza mais a inserção do sistema cooperativo de bicicletas do que o grupo do género feminino.

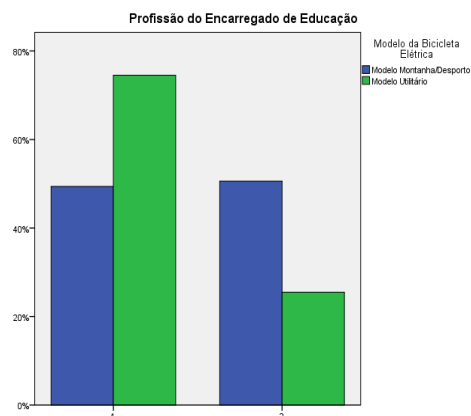
7. Grupo Socioeconómico

7.1. Regime de utilização da bicicleta elétrica



O grupo socioeconómico de baixo rendimento (correspondente à categoria 2 da figura acima) prefere utilizar a bicicleta convencional, enquanto os grupos socioeconómicos de alto e médio rendimento (correspondentes à categoria 1 da figura acima) prefere não optar por nenhum regime de utilização.

7.2. Modelos da bicicleta elétrica



De acordo com a figura anterior, o grupo socioeconómico de alto e médio rendimento (categoria 1) prefere o modelo utilitário, enquanto o grupo socioeconómico de baixo rendimento (categoria 2) tem preferência pelo modelo montanha/desporto.

Apêndice 2

Correlação entre as covariáveis

Correlação	Intercept (Disponibilidade do automóvel)	Custo	Comprimento da ciclovias	Novas TIC
Custo	0.99			
Comprimento da ciclovias	-0.22	-0.19		
Novas TIC	-0.14	-0.12	-0.01	
Tempo de viagem	0.09	0.17	-0.13	-0.03

Tabela 1: Correlação dos parâmetros do modelo base I

Correlação	Custo	Comprimento da ciclovias	Novas TIC
Comprimento da Ciclovias	0.42		
Novas TIC	0.21	-0.10	
Tempo de viagem	0.81	0.02	0.00

Tabela 2: Correlação dos parâmetros do modelo base II

Para que o modelo de regressão logística seja ajustado de forma eficiente, é preciso que todas as covariáveis que este apresenta não sejam correlacionadas entre si, caso contrário o modelo apresentaria uma forte multicolinearidade (forte correlação entre as covariáveis). Deste modo, na presença de multicolinearidade, a estimativa do impacto de uma covariável sobre a variável resposta, mantendo as restantes covariáveis constantes, tende a ser menos precisa do que caso as covariáveis sejam não correlacionadas. Por outro lado, as covariáveis colineares contém a mesma informação sobre a variável dependente (variável resposta). Se as medidas nominalmente diferentes quantificam o mesmo fenómeno, então estas são redundantes. Em alternativa, se as covariáveis apresentam diferentes escalas (ex: idade, altura, peso,...etc.) mas são fortemente correlacionadas entre si, então estas apresentam uma forte redundância no efeito sobre a variável resposta.

Anexo 1

Tabela 1: Legislação da Bicicleta Elétrica implementada em Portugal

Lei nº 22-A de 2007	Incorre numa reforma global na tributação automóvel, sendo aprovados o Código do Imposto sobre Veículos e o Código do Imposto Único de Circulação e abolidos o Imposto Automóvel, o Imposto Municipal sobre Veículos, o Imposto de Circulação e o Imposto de Camionagem.
Resolução do Conselho de Ministros nº 20/2009	Cria o Programa para a Mobilidade Elétrica em Portugal, dirigido por um gabinete constituído no âmbito do Ministério da Economia e da Inovação, que tem como objetivo a introdução e massificação da utilização do veículo elétrico.
Resolução do Conselho de Ministros nº 81/2009	Aprova metas e objetivos do Programa para a Mobilidade Elétrica, no que diz respeito à definição do Programa e do respetivo plano de trabalho, atividades e prazos, atores envolvidos e responsabilidades.
Decreto-Lei nº 39/2010	Estabelece o regime jurídico da mobilidade elétrica, aplicável à organização, acesso e exercício das atividades relativas à mobilidade elétrica, bem como as regras destinadas à criação de uma rede piloto de mobilidade elétrica.
Decreto-Lei nº 140/2010	Transpõe, para a ordem jurídica interna, a Diretiva nº 2009/33/CE, do Parlamento Europeu e do Conselho, de 23 de Abril, relativa à promoção de veículos de transporte rodoviário não poluentes e energeticamente eficientes. Estabelece, no âmbito da Estratégia Nacional da Energia 2020, o regime jurídico relativo à promoção da aquisição ou locação por entidades públicas de veículos de transporte rodoviário não poluentes e energeticamente eficientes.
Portaria nº 456/2011	Estabelece os requisitos técnicos e financeiros a que fica sujeita a atribuição de licença para o exercício da atividade de comercialização de eletricidade para a mobilidade elétrica, bem como algumas regras de procedimento aplicáveis à instrução do respetivo requerimento.
Portaria nº 468/2011	Estabelece os termos em que são concedidos os incentivos financeiros à aquisição de veículos novos exclusivamente elétricos previstos no artigo 38.º do Decreto -Lei n.º

	39/2010, de 26 de Abril.
Portaria nº 1201/2011	Estabelece os requisitos técnicos para o licenciamento da atividade de operação de postos de carregamento da rede de mobilidade elétrica.
Portaria nº 1202/2011	Estabelece os termos aplicáveis às licenças de utilização privativa do domínio público para a instalação de postos de carregamento de baterias de veículos elétricos em local público de acesso público.
Portaria nº 1232/2011	Fixa o valor das taxas devidas pela emissão das licenças de comercialização de eletricidade para a mobilidade elétrica e de operação de pontos de carregamento, bem como da taxa de inspeção devida pela realização de inspeções periódicas.
Portaria nº 173/2011	Estabelece as condições mínimas, os limites de capital e os riscos cobertos pelo seguro obrigatório de responsabilidade civil por danos causados no exercício da atividade de comercialização de eletricidade para a mobilidade elétrica.
Portaria nº 180/2011	Regula os termos e condições essenciais da remuneração da atividade de operação de pontos de carregamento: a) Nos pontos de carregamento normal de acesso público, b) Nos pontos de carregamento rápido. Define, ainda, o montante máximo que pode ser auferido pela atividade de manutenção de pontos de carregamento de acesso privativo em locais de estacionamento em prédios urbanos para fins residenciais.
Portaria nº 252/2011	Estabelece as normas técnicas para instalação e funcionamento de pontos de carregamento normal em edifícios e outras operações urbanísticas.
Regulamento nº 464/2011	Estabelece disposições aplicáveis ao exercício das atividades de mobilidade elétrica abrangidas pela regulação da ERSE.
Lei nº 64-B/2011	Estabelece a revogação dos incentivos financeiros na aquisição de veículos elétricos, nomeadamente: 1 - É revogado o capítulo V do Decreto-Lei n.º 39/2010, que cria e regulamenta os incentivos financeiros na aquisição de veículos exclusivamente elétricos; 2 - Os certificados de destruição emitidos nos termos do Decreto-Lei n.º 39/2010, que habilitam ao incentivo financeiro de 1500 euros, perdem a sua validade em 31 de Dezembro de 2011. São

	<p>eliminados os incentivos à aquisição de viaturas elétricas, que abrangia um incentivo financeiro e um incentivo à destruição de viaturas com mais de 10 anos, em cumprimento do compromisso assumido no Memorando de Entendimento celebrado pelo Estado Português com a União Europeia, o FMI e o BCE de avaliar os atuais instrumentos relacionados com a energia, incluindo os incentivos fiscais em matéria de eficiência energética, designadamente da avaliação do risco de sobreposição ou de inconsistência de instrumentos.</p>
Diretiva nº 1/2012	<p>Procede à alteração do Regulamento da Mobilidade Elétrica, Regulamento n.º 464/2011.</p>
Decreto-Lei nº 170/2012	<p>Procede à primeira alteração do Decreto-Lei n.º 39/2010 de 26 de Abril, que cria o regime jurídico da mobilidade elétrica.</p>
Declaração de Retificação nº 47/2012	<p>Retifica o Decreto-Lei n.º 170/2012 de 1 de Agosto, que procede à primeira alteração do Decreto-Lei n.º 39/2010 de 26 de Abril, que cria o regime jurídico da mobilidade elétrica.</p>
Despacho nº 115/2013	<p>Prorroga até à data de aprovação da nova estratégia de revisão do Programa para a Mobilidade Elétrica, ou até ao dia 30 de junho de 2013, consoante o que ocorra primeiro, a fase piloto do Programa para a Mobilidade Elétrica. Constitui, sob a orientação e nos termos a definir pelos Secretários de Estado do Empreendedorismo, Competitividade e Inovação e da Energia, uma equipa encarregada da revisão técnica do Programa para a Mobilidade Elétrica.</p>